APPLICATIONS OF QUANTILE REGRESSION TO ESTIMATION AND DETECTION OF SOME TAIL CHARACTERISTICS

BY

YA-HUI HSU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Statistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Professor Xuming He, Chair
Professor Roger Koenker
Assistant Professor Feng Liang
Professor Steve Portnoy

Abstract

The statistical inference based on the ordinary least squares regression is sub-optimal when the distributions are skewed or when the quantity of interest is the upper or lower tail of the distributions. For example, the changes in Total Sharp Scores (TSS), the primary measurements of the treatment effects on prevention of structural damage for rheumatoid arthritis, are nearly identical for most therapies for nearly 75% of the patient population, but the difference lies in the most challenging 25% of the patient population where a less effective treatment loses its efficacy, resulting in a heavy right tail in its distribution.

In the first part of the dissertation, we develop the Expected Shortfall (ES), the Covariate-adjusted Expected Shortfall (COVES), and the Generalized Covariate-adjusted Expected Shortfall (q.COVES) tests under the framework of quantile regression. Those tests focus specifically on one tail of the outcome distributions. The ES test applies to two-sample comparisons. The COVES test adjusts for covariates, and is shown to be valid for i.i.d (independent and identically distributed) error models or when the covariates have the same means across treatments. The q.COVES test generalizes the COVES test to more general models. We show the proposed tests can achieve a substantial sample size reduction over the conventional tests on mean effects.

The second part of the dissertation focuses on a popular measure of risk used by financial institutions, Value at Risk (VaR), defined as a quantile of the loss distribution of a portfolio within a given time period and a confidence level. Accurate VaR estimation can help financial institutions maintain appropriate capital levels to cover the risk from the corresponding portfolio. We use an MCMC strategy along with a block algorithm to perform Bayesian inference on the Conditional Autoregressive Value at Risk (CAViaR) models proposed by Engle and Manganelli (2004) based on quantile regression. Using the S&P 500 index as an example, we show that the proposed Bayesian approach adds value to the original estimation method of Engle and Manganelli in terms of both estimation and prediction.

To my lovely family.

Acknowledgments

Being a PhD student in the Department of Statistics at the University of Illinois Urbana-Champaign is a wonderful experience and a good investment to me. I want to thank many people who helped me make my dissertation possible.

I am so grateful to meet Professor Xuming He and to have an opportunity working on my dissertation with him. His knowledge on quantile regression and risk management, and his incredible ideas on reserach have been of great value to me. I am so lucky to have him as my advisor who has spent lots of time discussing my research, revising my dissertation, and advising on my career. I always understand the problems better and develop better problem-solving abilities after discussion with him.

I would also like to express my deepest gratitude to my other committee members: Professors Roger Koenker, Feng Liang, and Steve Portnoy. Their valuable suggestions guided me to explore my research further.

I would like to thank Dr. Mingxiu Hu, who offered me the motivation on one of my research topics and a valuable dataset. He also offered me an opportunity to work as an intern at Millennium: the Takeda Oncology company. The experience gained as an intern helped me find a job in the health and pharmaceutical industry after my PhD.

I am grateful to have met all the faculty, staff, and fellow students of the Department of Statistics at the University of Illinois Urbana-Champaign. Their support has been valuable to me. Special thanks to Maria Muyot, Adam Martinsek, Usha Dhar, Melissa Banks, Judy Whittington, Xingdong Feng, Zhi He, Yang Feng, Guixian Lin, Feng Hong, Grace Hong, Ji Yeon Yang, Ji Young Kim, and Yunwen Yang, for their help over the years.

Most important is the understanding and encouragement from my lovely family and relatives. Through this long journey of study, they always offer me ultimate love and support when I need them most. Finally, I appreciate the financial support from the National Science Foundation and National Institute of Health through grants to Professor Xuming He.

Table of Contents

List of Tables

List of Figures

Chapter 1

Expected Shortfall Test for Detection of Treatment Effects

1.1 Introduction

We consider the problem of testing the hypothesis of no treatment effect against a class of alternatives where the two outcome distributions differ only or mainly in the right tail. As demonstrated in some recent trials of rheumatoid arthritis therapies in van der Heijde et al. (2006) and Kremer et al. (2006), the changes in Total Sharp Scores (TSS), the primary measurements of the treatment effects on prevention of structural damage, are nearly identical for most therapies for nearly 75% of the patient population, but the difference lies in the most challenging 25% of the patient population where a less effective treatment loses its efficacy, resulting in a heavy right tail in its distribution. The two-sample t test or its regression counterpart in covariate-adjusted linear models is commonly used for detecting the treatment effects, but due to skewness and heavy-tails of the distributions, the test does not have satisfactory power. Nonparametric tests on the median differences, for example, would fare even worse in such cases, because the median differences are often negligible among those therapies. Some trials have considered the chi-square tests on the proportion of patients with little disease progression by dichotomizing TSS, but there has been no agreeable cutoff point for dichotomization. In fact, the power of the chi-square test depends rather critically on the cutoff point. In addition, it is difficult to perform the chi-square test when a covariate needs to be adjusted for.

We believe that a most natural test in this type of applications is to focus on the average in one tail, or the expected tail loss (aka expected shortfall, abbreviated as ES). In finance, this is also called the conditional value at risk (C-VaR) (Artzner et al., 1999), for measuring the risk of a portfolio. We average the changes in TSS in the upper tail. A treatment is said to be more effective if it has a smaller expected shortfall, where the expected shortfall is defined to be the conditional mean of the outcome (e.g., change in TSS) above the τ-th quantile. In this paper, τ will be taken to be a user specified value (e.g., 0.75), but a good choice of τ clearly depends on the area of applications.

In finance, the most relevant choices of τ fall above 0.90.

O'Brien (1998) proposes the generalized t-test and the generalized rank sum test to incorporate heterogeneous treatment effects. These tests have good power against location-scale changes, but are limited to two-sample comparisons without covariates. A main advantage of the ES test is that it generalizes to adjust for covariates, and the generalization will appear in the next chapter.

The remainder of this chapter is organized as follows. We start with a brief introduction to our motivating example on the TSS for rheumatoid arthritis therapies in the next section. Next, we propose the ES test based on the expected shortfalls and compare the proposed ES test with the t test and the χ^2 test in empirical power.

1.2 A Premier on Total Sharpe Scores

Rheumatoid arthritis (RA) is a chronic disabling disease that causes destruction of joint cartilage and erosion of adjacent bones. In RA clinical trials, TSS is used to measure the treatment effect of RA drugs on prevention of structural damage to the joints. It consists of two components, erosion score and score for joint space narrowing (JSN), which are obtained through examination of hand and/or feet joints with radiographic methods. The first description of TSS is given by Sharp et al. (1971), but TSS has been modified in later studies. The example presented in this section is based on van der Heijde's modification of TSS scoring system (van der Heijde, 2000), which is based on examination of 16 areas for erosions and 15 for joint space narrowing in each hand. The erosion score per joint ranges from 0 to 5 with 0 representing a normal condition and 5 the most severe disease, and thus the total erosion score ranges from 0 to 160 (16 areas by 2 hands by 5). The JSN score ranges from 0 to 4 per joint with higher score representing more severe disease, which leads to a range of 0 to 120 (15 areas by 2 hands by 4) for the total JSN score. Therefore, the range of TSS is 0 to 280. The primary interest is the change from baseline in TSS in one or two years.

The change in TSS has a highly skewed distribution under any known treatment. In the TEMPO trial (van der Heijde et al., 2006) comparing Methotrexate, Etanercept, and the combination therapy of Etanercept and Methotrexate, the three treatments are similarly effective for about 75% of the patients whose conditions improved or showed no or little progression from the baseline; see Figure 1.1. Medians for all three groups are around 0. Treatment differences come from the 25% of the patients with the most progressive diseases. In other words, the differences in treatment effects are not attributed to a location-scale change in the distributions. The distributions of clinical data from several other major RA trials (Kremer et al., 2006; Keystone et al., 2004; Lipsky et al., 2000) showed similar characteristics.

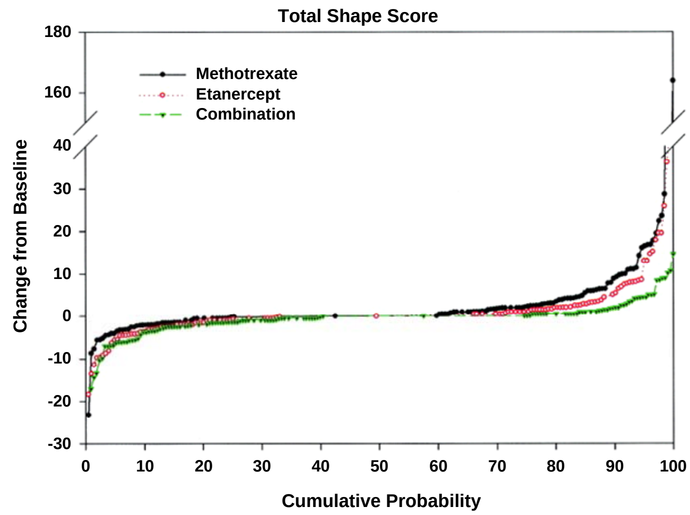


Figure 1.1: This figure, reproduced from van der Heijde et al. (2006), shows that the changes in TSS from the TEMPO trials differ mostly in the upper tails.

Later in this chapter, we use a recent observational study conducted at Brigham and Women's Hospital and sponsored by Millennium Pharmaceuticals Inc. and Biogen Idec as a basis for assessing the performance of the proposed test. We take 150 subjects in the study, who are under active treatment, and simulate a control group whose outcome distribution is chosen to mimic the treatment difference reported in other trials. For example, in the Adalimumab trial (Keystone et al., 2004), the variance of the treatment group (using the drug Adalimumab 20mg/kg) is about half of that in the control group (using the drug Methotrexate) with a mean difference of -1.9. In the Abatacept trial (Kremer et al., 2006), the variance in the Abatacept group is about one third of that in the control group. In our simulation studies, we use the ratio of variances between 2:1 and 3:1 between two treatment groups.

1.3 Expected Shortfall (ES)

We use a dummy variable D as treatment indicator, and Z as the outcome measure. For simplicity of notation, we consider D taking values 0 or 1, but the work generalizes readily for multiple treatments. Expected shortfall (ES), also called Conditional Value at Risk (C-VaR) (Artzner et al., 1999) or Conditional Tail Expectation (CTE) or Tail-VaR, is the expected loss given that the loss has already exceeded the τth quantile of the loss distribution for some $\tau \in (0, 1)$. Expected shortfall is a robust and coherent measure (Artzner et al., 1999) for risk exposure expressed as $\mathrm{ES}_\tau(F) = \frac{1}{1-\tau}\int_\tau^1 F^{-1}(u)du$ for all loss distribution function, F.

Let Z be the outcome measure for the treatment ($d = 1$) and control ($d = 0$) groups with distributions F_1 and F_0, respectively. The hypothesis corresponding to expected shortfalls is stated as

$$H_0 : \mathrm{ES}_\tau(F_0) = \mathrm{ES}_\tau(F_1) \text{ versus } H_1 : \mathrm{ES}_\tau(F_0) = \mathrm{ES}_\tau(F_1) + \delta, \text{ for some } \delta \neq 0,$$

where $\mathrm{ES}_\tau(F_1)$ and $\mathrm{ES}_\tau(F_0)$ are the expected shortfalls for the treatment group and the control group, respectively, and δ is the effect size.

Given a random sample Z for group d, the empirical expected shortfall is

$$\mathrm{ES}_\tau(d) = \sum_{D_i=d} w_{d,i} Z_{(i)},$$

where $Z_{(i)}$ is the ith order statistic of Z for group d, $n_d = \sum_i I(D_i = d)$, $w_{d,i} = (n_d - [n_d\tau])^{-1} I(i > [n_d\tau])$ are the weights for group d, with $[x]$ denoting the largest integer not to exceed x, and I denoting the indicator function. The quantity $\mathrm{ES}_\tau(d)$ is the average of the outcomes for group d that are above the τ-th quantile. The proposed ES test statistic is given as

$$T_\tau^{ES}(n_1, n_0) = \mathrm{ES}_\tau(1) - \mathrm{ES}_\tau(0). \tag{1.1}$$

The ES test statistic in (1.1) is a linear function of order statistics, which is recognized as an L-estimator (Lehmann, 1983). Stigler (1974) shows that the L-estimator is asymptotically normal under some conditions. By using Theorem 5.1 of Lehmann (1983, p.369), we can obtain the asymptotic mean and variance of $T_\tau^{ES}(n_1, n_0)$.

Theorem 1.3.1. *(Theorem 5.1 of Lehmann, 1983)*
Let $X_1, ..., X_n$ be iid over an interval $(a, b), -\infty \leq a < b \leq \infty$ according to a distribution F for which $E(X_i^2) < \infty$ and which possesses a density f with $0 < f(x)$ for all $a < x < b$
(i) Let $L_n = \frac{1}{n}\sum_{i=1}^{n} \lambda(\frac{i}{n+1})X_{(i)}$ where λ is a bounded function defined over $(0, 1)$ which is continuous a.e. (with respect to Lebesgue measure) and satisfies $\int_0^1 \lambda(t)dt = 1$. Let $\mu(F, \lambda) = \int_0^1 \lambda(u)F^{-1}(u)du$ and $\sigma^2(F, \lambda) = \int_0^1 A^2(t)dt - (\int_0^1 A(t)dt)^2$ where A is any function with derivative $A'(t) = \frac{\lambda(t)}{f[F^{-1}(t)]}$. Then the distribution of $\sqrt{n}(L_n - \mu(F, \lambda))/\sigma(F, \lambda)$ tends to the standard normal distribution $N(0, 1)$

as $n \to \infty$, provided $\sigma^2(F,\lambda) > 0$.

(ii) The result of part (i) remains valid if the density λ is replaced by λ_n, provided the sequence λ_n is uniformly bounded and $\lambda_n(t) \to \lambda(t)$ where λ is continuous a.e. and the convergence is uniform in a neighborhood of any continuity point of λ.

We quote the result from Theorem 1.3.1, and apply this to establish the asymptotic normality of the test statistic for (1.1) as $n_1, n_0 \to \infty$.

Theorem 1.3.2. *Let (Z_i, D_i) with $D_i = 1$ for $i = 1, \cdots, n_1$ and $D_i = 0$ for $i = n_1+1, \cdots, n_1+n_0$, Z be a combination of two random samples according to distributions F_1 and F_0, respectively, with finite second moments, then*

$$(T_\tau^{ES}(n_1,n_0) - \mu^{ES})/\sigma_{n_1,n_0}^{ES} \xrightarrow{\mathcal{D}} N(0,1), \text{ as } n_1,\ n_0 \to \infty$$

with asymptotic mean

$$\mu^{ES} = ES_\tau(F_1) - ES_\tau(F_0),$$

and asymptotic variance

$$(\sigma_{n_1,n_0}^{ES})^2 = \sigma^2(F_1,\tau)/n_1 + \sigma^2(F_0,\tau)/n_0,$$

where

$$\begin{aligned} ES_\tau(F_d) &= \frac{1}{1-\tau}\int_\tau^1 F_d^{-1}(u)du, \\ \sigma^2(F_d,\tau) &= \frac{1}{(1-\tau)^2}\int_\tau^1 (F_d^{-1}(t) - F_d^{-1}(\tau))^2 dt - \Big(\frac{1}{1-\tau}\int_\tau^1 (F_d^{-1}(t) - F_d^{-1}(\tau))dt\Big)^2,\ d = 0, 1. \end{aligned}$$

Proof of Theorem 1.3.2 Let $\lambda(t) = \frac{1}{1-\tau} I(\tau < t < 1)$, where $\lambda(t)$ is bounded and continuous a.e. over $(0,1)$ and satisfies $\int_0^1 \lambda(t)dt = 1$. According to Theorem 1.3.1, $\mathrm{ES}_\tau(d)$ is asymptotically normal with mean $\mu_d = \frac{1}{1-\tau}\int_\tau^1 F_d^{-1}(u)du$ and variance $\sigma_d^2 = \sigma^2(F_d,\tau)/n_d$, where $d = 0,\ 1$.

To compute $\sigma^2(F_d, \tau)$, we follows Theorem 1.3.1 to get $A'_{Z_d}(t)$, and $A_{Z_d}(t)$ first.

$$\begin{aligned}
A'_{Z_d}(t) &= \frac{1}{1-\tau}\frac{1}{f_d(F_d^{-1}(u))}I(\tau < t < 1), \\
A_{Z_d}(t) &= \int_\tau^t \frac{1}{1-\tau}\frac{1}{f_d(F_d^{-1}(u))}du = \frac{1}{1-\tau}(F_d^{-1}(t) - F_d^{-1}(\tau))I(\tau < t < 1), \\
\sigma^2(F_d, \tau) &= \int_0^1 A^2_{Z_d}(t)dt - (\int_0^1 A_{Z_d}(t)dt)^2 \\
&= \frac{1}{(1-\tau)^2}\int_\tau^1 (F_d^{-1}(t) - F_d^{-1}(\tau))^2 dt - \Big(\frac{1}{1-\tau}\int_\tau^1 (F_d^{-1}(t) - F_d^{-1}(\tau))dt\Big)^2.
\end{aligned}$$

Since Z are independent for $d = 0$ and $d = 1$, $T_\tau^{ES}(n_1, n_0)$ is asymptotically normal with the mean and the variance specified in this theorem. $\diamond$

Lemma 1.3.1. *Let Z be a combination of two random samples according to distributions F_1 and F_0, respectively, with finite second moments. We have*

$$\begin{aligned}
&Z_{([n_d\tau])} = F_d^{-1}(\tau) + o_p(1), \\
&(Z_{([n_d\tau])})^2 = (F_d^{-1}(\tau))^2 + o_p(1), \\
&\frac{1}{(1-\tau)n_d}\sum_{i>[n_d\tau],\ D_i=d} Z_{(i)} = \frac{1}{1-\tau}\int_\tau^1 F_d^{-1}(t)dt + o_p(1), \\
&\frac{1}{(1-\tau)n_d}\sum_{i>[n_d\tau],\ D_i=d} (Z_{(i)})^2 = \frac{1}{1-\tau}\int_\tau^1 (F_d^{-1}(t))^2 dt + o_p(1).
\end{aligned}$$

Proof of Lemma 1.3.1 According to Serfling (2002),

$$\sqrt{n_d}(Z_{([n_d\tau])} - F_d^{-1}(\tau)) \xrightarrow{\mathcal{D}} N(0,\ \frac{\tau(1-\tau)}{\{f_d(F_d^{-1}(\tau))\}^2}),\ \text{as } n_d \to \infty,$$

$Z_{([n_d\tau])} = F_d^{-1}(\tau) + O_p(n_d^{-1/2}) = F_d^{-1}(\tau) + o_p(1)$, and $(Z_{([n_d\tau])})^2 = (F_d^{-1}(\tau))^2 + o_p(1)$. Besides,

$$\begin{aligned}
\frac{1}{(1-\tau)n_d}\sum_{i>[n_d\tau],\ D_i=d} Z_{(i)} &= \frac{1}{1-\tau}[\frac{1}{n_d}\sum_{D_i=d}\{Z_i I(Z_i \geq F_d^{-1}(\tau))\}] + o_p(1) \\
&= \frac{1}{1-\tau}E\big(Z_i I(Z_i \geq F_d^{-1}(\tau),\ D_i = d)\big) + o_p(1) \\
&= \frac{1}{1-\tau}\int_{F_d^{-1}(\tau)}^\infty z dF_d(z) + o_p(1) \\
&= \frac{1}{1-\tau}\int_\tau^1 F_d^{-1}(t)dt + o_p(1).
\end{aligned}$$

Similarly, we can prove

$$\frac{1}{(1-\tau)n_d}\sum_{i>[n_d\tau],\ D_i=d}(Z_{(i)})^2 = \frac{1}{1-\tau}\int_{\tau}^{1}(F_d^{-1}(t))^2dt + o_p(1). \ \diamond$$

Theorem 1.3.3. *Let Z be a combination of two random samples according to distributions F_1 and F_0, respectively, with finite second moments. Under the null hypothesis that $\mu^{ES}=0$, we have*

$$T_{\tau}^{ES}(n_1,n_0)/s_{n_1,n_0}^{ES} \xrightarrow{\mathcal{D}} N(0,1), \ \text{as } n_1, \ n_0 \to \infty,$$

where

$$(s_{n_1,n_0}^{ES})^2 = V_1^{ES}/n_1 + V_0^{ES}/n_0,$$

and

$$V_d^{ES} = (1-\tau)^{-2}n_d^{-1}\sum_{i>[n_d\tau],\ D_i=d}\{Z_{(i)}-Z_{([n_d\tau])}\}^2 - \left[\{(1-\tau)n_d\}^{-1}\sum_{i>[n_d\tau],\ D_i=d}\{Z_{(i)}-Z_{([n_d\tau])}\}\right]^2$$

is the consistent estimator for $\sigma^2(F_d,\tau)$.

Proof of Theorem 1.3.3 According to Lemma 1.3.1,

$$\frac{1}{(1-\tau)n_d}\sum_{i>[n_d\tau],\ D_i=d}\{Z_{(i)}-Z_{([n_d\tau])}\} \text{ and } \frac{1}{(1-\tau)n_d}\sum_{i>[n_d\tau],\ D_i=d}\{Z_{(i)}-Z_{([n_d\tau])}\}^2$$

are consistent estimators for

$$\frac{1}{1-\tau}\int_{\tau}^{1}(F_d^{-1}(t)-F_d^{-1}(\tau))dt \text{ and } \frac{1}{1-\tau}\int_{\tau}^{1}(F_d^{-1}(t)-F_d^{-1}(\tau))^2dt, \text{ respectively.}$$

Finally, according to Theorem 1.3.2, the proof is complete. $\diamond$

1.4 Simulations

Suppose that Z is the response variable under two treatments according to distributions F_1 and F_0. We consider the problem of testing the null hypothesis that $\text{ES}_\tau(F_0) = \text{ES}_\tau(F_1)$ versus two-sided alternative hypothesis. Four tests are considered to detect the treatment effect, where the difference may lie in the upper tail of the distributions. The $0.5^{\text{th}}/0.75^{\text{th}}$ ES test is based on the difference in the expected loss above the $0.5^{\text{th}}/0.75^{\text{th}}$ quantile, the χ^2 test is based on a 2 by 2 table, where success is defined as the measurement below a given cutoff point, and the t test is based on the difference in the sample means. Since the cutoff point is critical in determining the power, we choose it to be median over one standard deviation of the treatment group as van der Heijde et al. (2006) suggested.

As in the Adalimumab and Abatacept trials (Keystone et al., 2004; Kremer et al., 2006), in which the variance of the treatment group is around half or one third of that of the control group, the treatment group usually has a smaller variance, and a similar or larger sample size than the control group does. So in the simulation study, we take the ratio of sample size (n_1/n_0) to be 1 or 2, where n_1 and n_0 are the sample sizes for the treatment and control group, respectively. In this section, we report some power studies of the proposed test based on Monte Carlo simulations.

1.4.1 Some Simulation Studies

We consider data generated from three studies in the analysis.

- *Study 1, normal models:* $Z_i = 5 + \{1 + \eta I(e_i > 0)I(D_i = 0)\}e_i$, where $e_i \sim N(0, 1)$, and η is either 0 (under the null hypothesis) or 1.35 (under the alternative hypothesis).

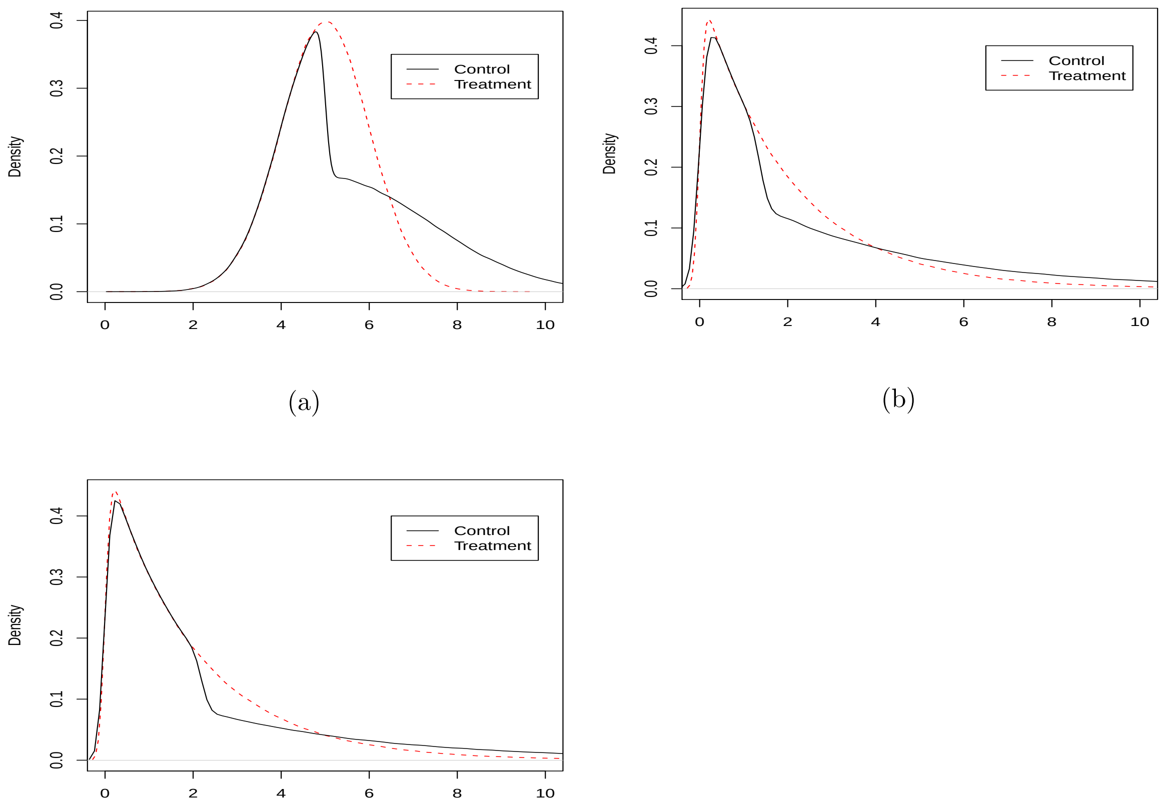
- *Study 2, Chi-square models with agreement up to the median:* $Z_i = e_i + \eta(e_i - med)I(e_i > med)I(D_i = 0)$, where $e_i \sim \chi^2(\nu)$, degree of freedom $(\nu) = 2$, median (med) ≈ 1.4, and η is either 0 (under the null hypothesis) or 0.85 (under the alternative hypothesis).

- *Study 3, Chi-square models with agreement up to a higher quantile:* $Z_i = e_i + \eta(e_i - 2.2)I(e_i > 2.2)I(D_i = 0)$, where $e_i \sim \chi^2(\nu)$, degree of freedom $(\nu) = 2$, and η is either 0 (under the null hypothesis) or 1.05 (under the alternative hypothesis).

Clearly, the control group ($d = 0$) has a heavier right tail. Figure 1.2 shows the density function for two groups in each study. When η equals to the value under the alternative hypothesis, the error variance of the control group ($d = 0$) is about triple that of the treatment group ($d = 1$) in each study. Table 1.1 summarizes the differences of the two groups under the alternative hypothesis in each study.

Type I Error and Histogram of p-values

A total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 50$ are generated for Z under the null hypothesis (with $\eta = 0$). We show the histograms of the p-values computed from all tests for three studies in Figure 1.3 and Figure 1.4. The Type I errors stay around the nominal level 5%, but



(a) (b)

(c)

Figure 1.2: Density functions of the treatment outcomes for the control group and treatment group, respectively. The upper tail of the control group is heavier than the treatment group. (a) Study 1. (b) Study 2. (c) Study 3.

the p-values from the χ^2 test deviates a little from the uniform distribution, which means the χ^2 approximation is not very accurate. The standard error of the Type I error estimate given under each figure is around 0.2% except for the χ^2 test.

Power Function

We compare power functions for these four tests, and then provide the sample sizes needed for each test to reach power 0.9. A total of $M = 10,000$ Monte Carlo samples are generated for Z under the alternative hypothesis (with $\eta = 1.35, 0.85$, and 1.05, respectively for Studies 1-3). The grid is 10 for sample size, and we use smoothing splines to obtain power curves. The cutoff point in the χ^2 test is 6 for Study 1, and 3.4 for Studies 2 and 3. Figure 1.5 shows that the 0.5th/0.75th ES test is more powerful than the χ^2 test and the t test. Under the alternative hypothesis, the ES test is indeed better for detecting a difference than the χ^2 and the t tests. The results from Figure 1.5 and

Table 1.1: Difference of the two groups ($d = 0$ vs $d = 1$) under the alternative hypothesis in each study, with the last column for the ratio of error variances.

τ	0.5	0.6	0.7	0.75	0.8	0.9	mean	var ratio
Study 1 ($\eta = 1.35$)	0	0.34	0.70	0.91	1.13	1.72	0.54	**2.97**
Study 2 ($\eta = 0.85$)	0	0.37	0.85	1.16	1.54	2.73	0.84	**2.96**
Study 3 ($\eta = 1.05$)	0	0	0.22	0.60	1.07	2.52	0.70	**3.07**

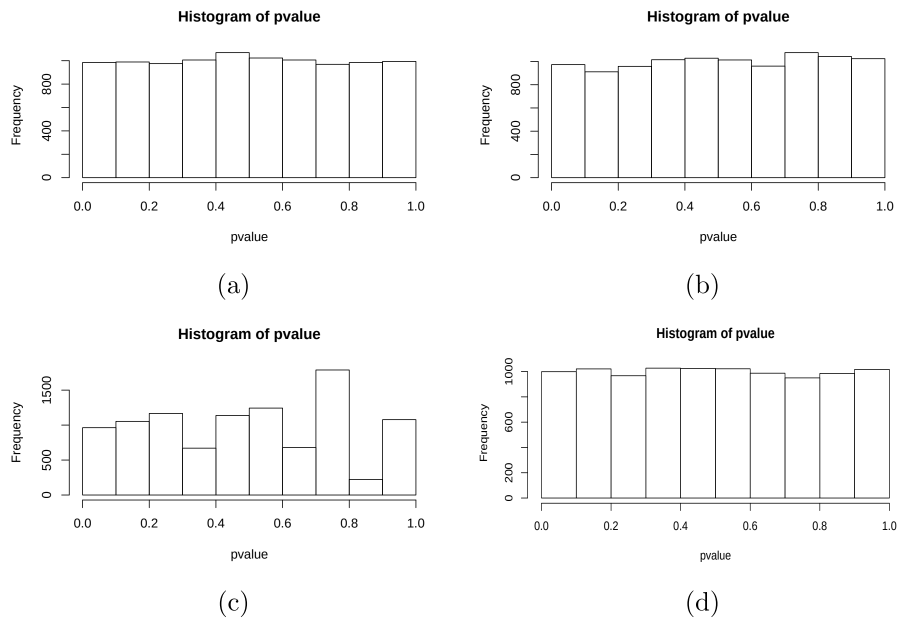
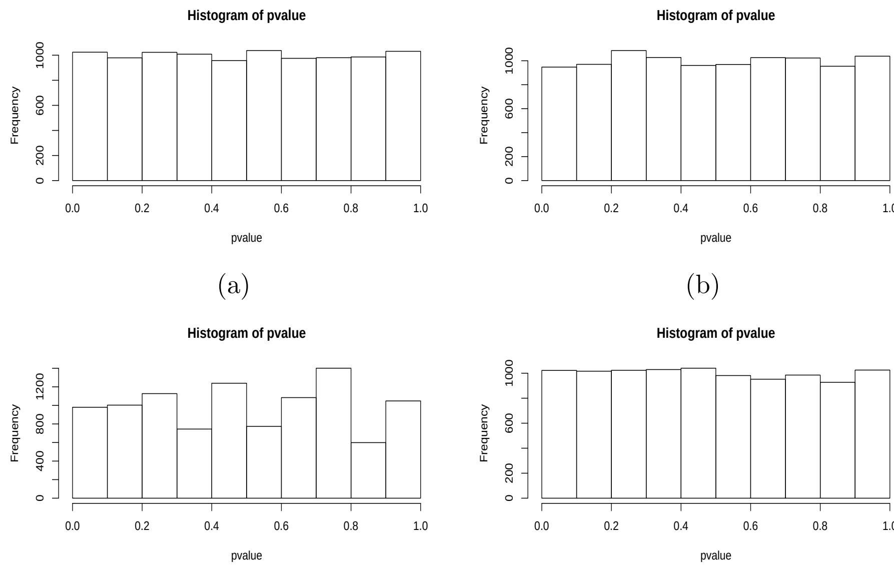


Figure 1.3: Histograms of p-values for Study 1. (a) 0.5th ES test, Type I error = 0.0502. (b) 0.75th ES test, Type I error = 0.0512. (c) χ^2 test, Type I error = 0.0522. (d) t test, Type I error = 0.0496.

Table 1.2 clearly show the superiority of the ES test. In Study 1, if we use the t test, we need to have a sample size twice as large compared to the ES test.

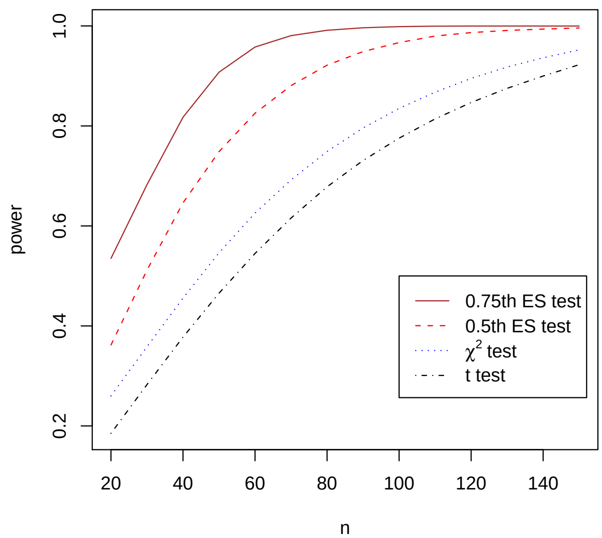


(c) (d)

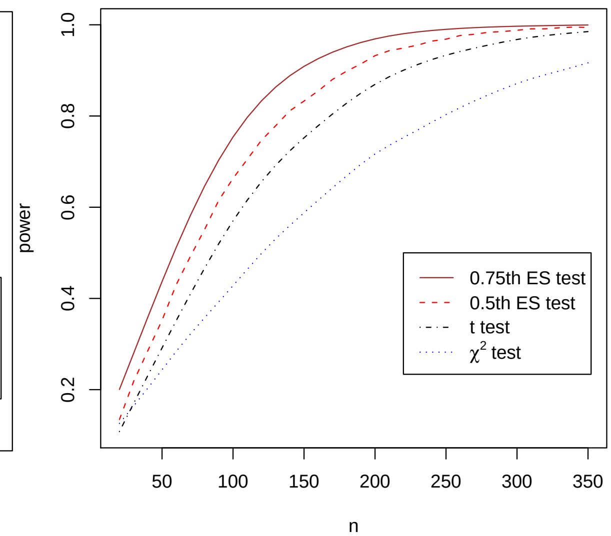
Figure 1.4: Histograms of p-values for Study 2 and Study 3. (a) 0.5th ES test, Type I error = 0.0498. (b) 0.75th ES test, Type I error = 0.0483. (c) χ^2 test, Type I error = 0.0512. (d) t test, Type I error = 0.0499.

Table 1.2: Comparisons of four tests. (a) Type I error at sample size $(n_1,\ n_0) = (50, 50)$. The values in the table are in percentage. (b) Sample size $(n_1,\ n_0)$ needed to reach power 0.9 under the alternative hypothesis. In each study, the first row uses $n_1 = n_0$, and the second row uses $n_1 = 2n_0$.

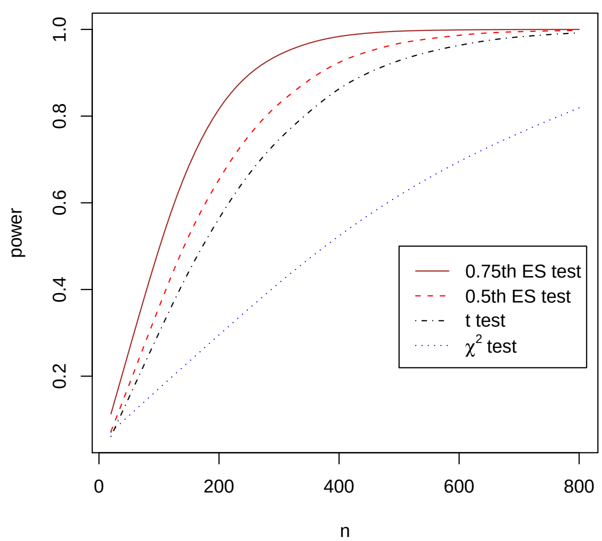
	0.5th ES test		0.75th ES test		χ^2 test		t test	
Study	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1	5.02	(76, 76)	5.12	(50, 50)	5.22	(121, 121)	4.96	(141, 141)
		(136, 68)		(92, 46)		(180, 90)		(242, 121)
2	4.98	(180, 180)	4.83	(147, 147)	5.12	(331, 331)	4.99	(221, 221)
		(308, 154)		(260, 130)		(490, 245)		(372, 186)
3	4.98	(274, 274)	4.83	(195, 195)	5.12	(817, 817)	4.99	(335, 335)
		(490, 245)		(342, 171)		(1214, 607)		(560, 280)



(a)



(b)



(c)

Figure 1.5: Statistical powers of four tests as functions of sample size $n_1 = n_0 = n$. (a) Study 1. (b) Study 2. (c) Study 3.

1.4.2 Targeted Study on TSS

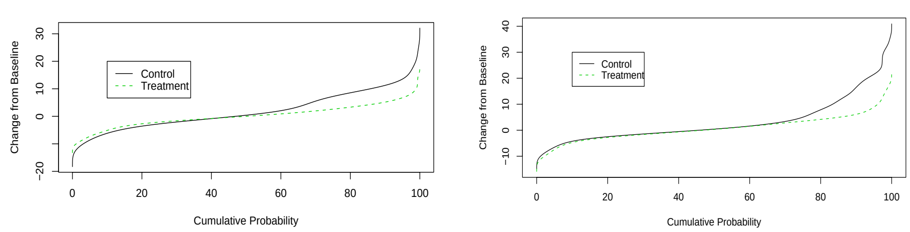
Two scenarios (with variance ratio 2 and 3, respectively) are constructed based on the data we obtained from a recent study on an undisclosed therapy to treat RA at the Brigham and Women's Hospital in Boston. Since the cutoff point is critical to determining the power, we choose the cutoff point to be one standard deviation of the treatment group as van der Heijde et al. (2006) suggested. One standard deviation is around 5.84 for this study. When the group differences occur at one tail of the distributions, we find that the proposed ES test greatly outperforms the t test.

TSS Data with Agreement up to certain Quantiles

We use the empirical distributions of the TSS changes of 150 patients in the Brigham and Women's Hospital study as the underlying distribution for the group $d = 1$. If this distribution is denoted as F (taken to be the empirical distribution from the data), the outcome measurement for the control group (with $d = 0$) will be simulated according to the following two scenarios.

- *Scenario 1, agreement up to 0.65th quantile:* $Z = F^{-1}(u) + 8|u - 0.65|^{1/4} I(u > 0.65)$,
- *Scenario 2, agreement up to 0.75th quantile:* $Z_i = F^{-1}(u) + 61|u - 0.75| I(u > 0.75)$,

where u is a uniform random number in (0, 1). Clearly, the control group ($d = 0$) has a heavier right tail. In fact, the variance of the control group ($d = 0$) is about twice and triple that of the treatment group ($d = 1$) under theses two scenarios, respectively. Figure 1.6 and Table 1.3 summarize the differences of the two groups.



(a) (b)

Figure 1.6: Cumulative probability distribution of the TSS change shows that the groups differ mostly in the upper trials. (a) Scenario 1. (b) Scenario 2.

Type I Error and Histogram of p-values

Histograms of the p-values at a total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 100$ from these tests for these two scenarios are shown in Figure 1.7. From Figure 1.7, we find that the

Table 1.3: Differences in the τ-th quantiles and in the mean, with the last column as the ratio of the variances between the control group ($d = 0$) and the treatment group ($d = 1$):

τ	0.5	0.6	0.7	0.75	0.8	0.9	0.99	mean	var ratio
Scenario 1	0	0	3.72	4.53	4.96	5.64	6.02	1.74	**2.03**
Scenario 2	0	0	0.03	1.14	3.24	9.08	14.58	1.99	**3.02**

results are similar to Section 1.4.1.

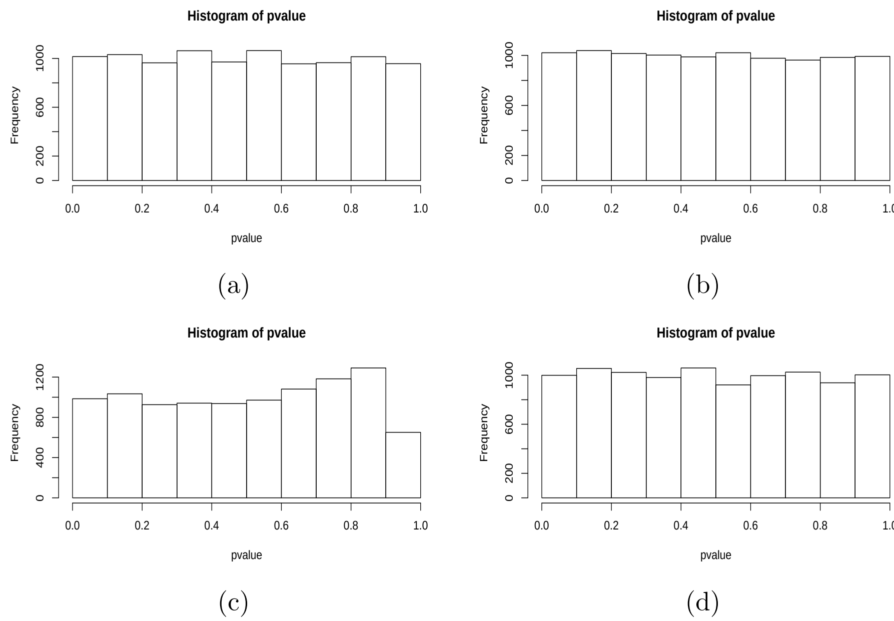
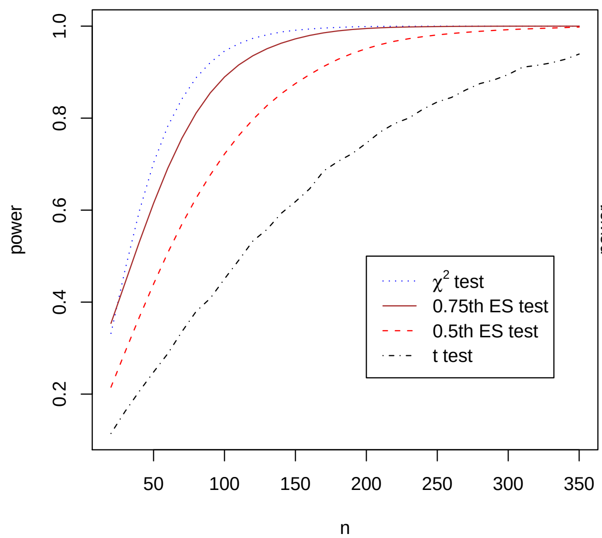


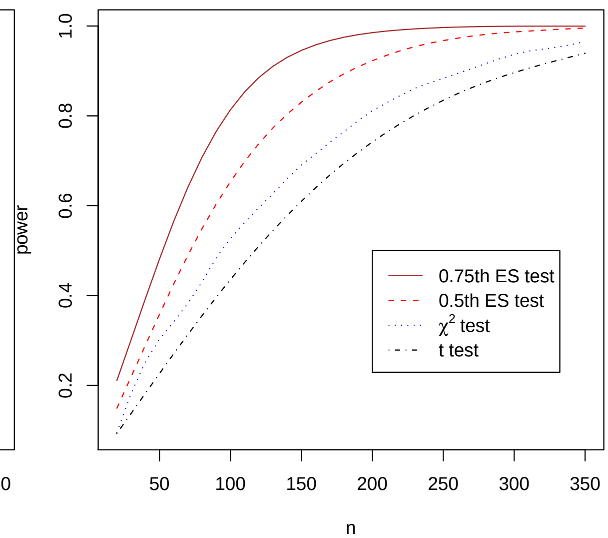
Figure 1.7: Histograms of p-values for both scenarios (a) 0.5th ES test, Type I error = 0.0500. (b) 0.75th ES test, Type I error = 0.0502. (c) χ^2 test, Type I error = 0.0491. (d) t test, Type I error = 0.0501.

Power Function

A total of $M = 10,000$ Monte Carlo samples are generated for Z. The power functions for all tests in each scenario are shown in Figure 1.8 with sample sizes up to $n_1 = n_0 = 350$. The grid is 10 for sample size, and we use smoothing splines to obtain power curves. Table 1.4 reports the Type I errors at the sample size of $n_1 = n_0 = 100$. It also reports the sample sizes needed to reach power of 0.90 in each scenario under two design conditions: $n_1 = n_0$ and $n_1 = 2n_0$, respectively. The results clearly show that both the ES test and the χ^2 test outperform the t test in these two scenarios. Besides, the ES test outperforms the χ^2 test in Scenario 2 but not in Scenario 1.



(a)



(b)

Figure 1.8: Statistical powers of four tests in reference with the TSS data with agreement up to the (a) 0.65th quantile and (b) 0.75th quantile, as functions of sample size $n_1 = n_0 = n$.

Table 1.4: Comparisons of four tests. (a) Type I error at sample size $(n_1,\ n_0) = (100, 100)$. The values in the table are in percentage. (b) Sample size $(n_1,\ n_0)$ needed to reach power 0.9 under the alternative hypothesis. In each study, the first row uses $n_1 = n_0$, and the second row uses $n_1 = 2n_0$.

	0.5th ES tets		0.75th ES test		χ^2 test		t test	
Scenario	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1	5.00	(170, 170)	5.02	(110, 110)	4.91	(85, 85)	5.01	(310, 310)
		(276, 138)		(158, 79)		(124, 62)		(508, 254)
2	5.00	(188, 188)	5.02	(130, 130)	4.91	(270, 270)	5.01	(310, 310)
		(322, 161)		(230, 115)		(390, 195)		(520, 260)

Chapter 2

Covariate-adjusted Expected Shortfall Test

2.1 Test Based on Covariate-adjusted Expected Shortfall (COVES)

Some covariates, C, may affect the treatment outcomes; therefore, taking covariates into consideration seems to be necessary in the analysis of treatment effects.

We use a dummy variable D as the treatment indicator, C as the covariates of interest, and Z as the outcome measure. For simplicity of notation, we consider D taking values 0 or 1, but the work generalizes readily for multiple treatments. Since the difference between two treatment groups may lie in the upper tail of the distributions, we model the τ-th quantile of Z given (D, C) as

$$Q_Z(\tau|D, C) = \alpha(\tau) + D\delta(\tau) + C\gamma^T(\tau) = X\beta(\tau), \tag{2.1}$$

where the coefficients α, δ, and γ are τ-specific, $X = (\mathbf{1}, D, C)$, and $\beta(\tau) = (\alpha(\tau), \delta(\tau), \gamma^T(\tau))^T$. In this chapter, we use $\tau = 0.5$ and 0.75 for simulation studies, and refer to Koenker (2005) for details on the linear regression quantile specification.

Given data (Z_i, D_i, C_i) with $D_i = 1$ for $i = 1, \cdots, n_1$ and $D_i = 0$ for $i = n_1 + 1, \cdots, n_1 + n_0$, we obtain the regression quantile coefficient $\hat{\alpha}(\tau)$, $\hat{\delta}(\tau)$, and $\hat{\gamma}(\tau)$ by quantile regression (Koenker and Bassett, 1978). Then, let $\hat{e}_i(\tau) = Z_i - \hat{\alpha}(\tau) - D_i\hat{\delta}(\tau) - C_i^T\hat{\gamma}(\tau)$ as the residuals from the τ-th regression quantile. By contrast, we also write $e_i(\tau) = Z_i - \alpha(\tau) - D_i\delta(\tau) - C_i^T\gamma(\tau)$, which has zero as the τ-th conditional quantile given (D_i, C_i) due to (2.1).

The outcome, Z_i, may be affected by both D_i and C_i. To focus on the treatment effect, we need to adjust the covariate effect. Let $Y_i = Z_i - C_i^T\hat{\gamma}(\tau)$ be the covariate-adjusted outcome, and define the empirical covariate-adjusted expected shortfall for the two groups as

$$\text{COVES}_\tau(d) = \sum_{D_i = d} w_{d,i} Y_i, \quad d = 0, 1,$$

where $w_{d,i} = S_d^{-1} I(\hat{e}_i(\tau) > 0)$ and $S_d = \sum_{D_i=d} I(\hat{e}_i(\tau) > 0)$. The quantity $\text{COVES}_\tau(d)$ is the average of the covariate-adjusted outcomes for group d that are above the τ-th quantile.

The proposed COVES test statistic for the hypothesis of no difference between the two treatment groups is given as

$$T_\tau^{COVES}(n_1, n_0) = \text{COVES}_\tau(1) - \text{COVES}_\tau(0). \tag{2.2}$$

Note that, when covariates of interest, C, are excluded, the COVES test is reduced to the ES test.

Let $\bar{C}_\tau(d)$ and $\bar{e}_\tau(d)$ be the average of C_i and e_i in group d that are above the τ-th regression quantile, i.e.,

$$\begin{aligned}
\bar{C}_\tau(d) &= S_d^{-1} \sum_{Di=d} \{C_i I(\hat{e}_i(\tau) > 0)\}, \\
\bar{e}_\tau(d) &= S_d^{-1} \sum_{Di=d} \{(Z_i - \alpha(\tau) - D_i\delta(\tau) - C_i^T \gamma(\tau)) I(\hat{e}_i(\tau) > 0)\}.
\end{aligned}$$

Then, the test statistic (2.2) can be written as

$$T_\tau^{COVES}(n_1, n_0) = \delta(\tau) - (\bar{C}_\tau(1) - \bar{C}_\tau(0))^T (\hat{\gamma}(\tau) - \gamma(\tau)) + (\bar{e}_\tau(1) - \bar{e}_\tau(0)), \tag{2.3}$$

which makes it relatively easy to establish the asymptotic normality of the test statistic as $n_1, n_0 \to \infty$.

To estimate the variance of $T_\tau^{COVES}(n_1, n_0)$, let $n_d = \sum_i I(D_i = d)$, f_i be the conditional density function of e_i given (D_i, C_i) evaluated at 0, and

$$C_i^* = C_i - n_d^{-1} \sum_i \{C_i I(D_i = d)\},$$

as the orthogonal components C relative to the treatment groups. In more general problems, we can obtain C^* by the Gram-Schmidt orthorgonalization of the design matrix. Furthermore, let

$$\begin{aligned}
V &= (n_1 + n_0)^{-1} \sum_i \{\hat{e}_i^2(\tau) I(\hat{e}_i(\tau) > 0)\} - \left[(n_1 + n_0)^{-1} \sum_i \{\hat{e}_i(\tau) I(\hat{e}_i(\tau) > 0)\}\right]^2, \qquad (2.4) \\
U_f &= \sum_i (f_i C_i^* C_i^{*T}), \quad \text{and} \\
s_{n_1,n_0}^2 &= (1-\tau)^{-2} V (n_1^{-1} + n_0^{-1}) + \tau(1-\tau)(\bar{C}_\tau(1) - \bar{C}_\tau(0))^T U_f^{-1} \{\sum_i C_i^* C_i^{*T}\} U_f^{-1} (\bar{C}_\tau(1) - \bar{C}_\tau(0)).
\end{aligned}$$

Lemma 2.1.1. *If $\{(Z_i, D_i, C_i)\}$ is a random sample satisfying (2.1), $\lim_{n_1,\, n_0 \to \infty} (n_1 + n_0)^{-1} U_f$ exists, $E \parallel C_i \parallel_1^3 < \infty$, and f_i are uniformly bounded away from 0 and infinity ($L < f_i < M$, $\forall i$), then we* *have the Bahadur representation on* $\hat{\gamma}(\tau)$

$$\hat{\gamma}(\tau) - \gamma(\tau) = U_f^{-1} \sum_i \{C_i^* \psi_\tau(e_i(\tau))\} + o_p((n_1 + n_0)^{-1/2}),$$

and the representation on $\bar{e}_\tau(d)$

$$\bar{e}_\tau(d) = \{\sum_{D_i = d} I(e_i(\tau) > 0)\}^{-1} \sum_{D_i = d} \{e_i(\tau) I(e_i(\tau) > 0)\} + o_p((n_1 + n_0)^{-1/2}),$$

where $U_f = \sum_i (f_i C_i^* C_i^{*T})$, f_i *is the conditional density function of* e_i *given* (D_i, C_i) *evaluated at 0, and* $C_i^* = C_i - n_d^{-1} \sum_i \{C_i I(D_i = d)\}$.

Proof of Lemma 2.1.1 The Bahadur representation for the quantile regression estimator, $\hat{\beta}$, is (Koenker, 2005)

$$\hat{\beta}(\tau) - \beta(\tau) \quad = \quad D_\beta^{-1} (n_1 + n_0)^{-1} \sum_i x_i \psi_\tau(e_i(\tau)) + (n_1 + n_0)^{-1/2} R_{n_1 + n_0},$$

where $\hat{\beta}(\tau) = (\hat{\alpha}(\tau), \hat{\delta}(\tau), \hat{\gamma}^T(\tau))^T$, $\beta(\tau) = (\alpha(\tau), \delta(\tau), \gamma^T(\tau))^T$, $x_i = (1, D_i, C_i^T)^T$, $\psi_\tau(e_i(\tau)) = \tau - I(e_i(\tau) < 0)$, $D_\beta = \lim_{n_1,\, n_0 \to \infty} (n_1 + n_0)^{-1} \sum_i f_i x_i x_i^T$, and $R_{n_1+n_0} = o_p(1)$.

For $X = (\mathbf{1}, D, C)$, we perform the Gram-Schmidt orthogonalization to get an orthogonal design matrix $X^* = (\mathbf{1}, D^*, C^*)$, where $D_i^* = D_i - \bar{D}_0$, $C_i^* = C_i - n_d^{-1} \sum_i C_i I(D_i = d)$, $\bar{D}_0$ is the overall mean for D, and $n_d = \sum_i I(D_i = d)$. Substituting the orthogonal design matrix X^* for X into the expression of D_β, we can get the diagonal matrix D_β^*. Let D_γ denote the right-bottom corner of D_β^*. Then,

$$\begin{aligned}
\hat{\gamma}(\tau) - \gamma(\tau) \quad &= \quad D_\gamma^{-1} (n_1 + n_0)^{-1} \sum_i \{C_i^* \psi_\tau(e_i(\tau))\} + o_p((n_1 + n_0)^{-1/2}) \\
&= \quad [\{(n_1 + n_0)^{-1} U_f\}^{-1} + o_p(1)](n_1 + n_0)^{-1} \sum_i \{C_i^* \psi_\tau(e_i(\tau))\} + o_p((n_1 + n_0)^{-1/2}) \\
&= \quad U_f^{-1} \sum_i C_i^* \psi_\tau(e_i(\tau)) + o_p(1)[(n_1 + n_0)^{-1} \sum_i \{C_i^* \psi_\tau(e_i(\tau))\}] + o_p((n_1 + n_0)^{-1/2}) \\
&= \quad U_f^{-1} \sum_i C_i^* \psi_\tau(e_i(\tau)) + o_p((n_1 + n_0)^{-1/2}),
\end{aligned}$$

where $U_f = \sum_i (f_i C_i^* C_i^{*T})$, f_i is the conditional density function of e_i given (D_i, C_i) evaluated at 0, and the last equality follows from the Central Limit Theorem for $\sum_i \{C_i^* \psi_\tau(e_i(\tau))\}$. $\diamond$

First, we prove the second equation in Lemma 2.1.1, which can be rewritten as

$$\{\sum_{D_i=d} I(\hat{e}_i(\tau)>0)\}^{-1}\sum_{D_i=d}\{e_i(\tau)I(\hat{e}_i(\tau)>0)\}-\{n_d(1-\tau)\}^{-1}\sum_{D_i=d}\{e_i(\tau)I(e_i(\tau)>0)\}=o_p((n_1+n_0)^{-1/2}).$$

We do this by verifying the following two statements.

$$n_d^{-1}\sum_{D_i=d} I(\hat{e}_i(\tau)>0)=1-\tau+o_p((n_1+n_0)^{-1/2}), \quad (2.5)$$

$$n_d^{-1}\big[\sum_{D_i=d} e_i(\tau)\{I(\hat{e}_i(\tau)>0)-I(e_i(\tau)>0)\}\big]=o_p((n_1+n_0)^{-1/2}). \quad (2.6)$$

By the second inequality in Corollary 2.1 of Koenker (2005), we have

$$n_d^{-1}\sum_{D_i=d} I(\hat{e}_i(\tau)>0)\le 1-\tau\le n_d^{-1}\sum_{D_i=d} I(\hat{e}_i(\tau)>0)+n_d^{-1}p,$$

and thus,

$$n_d^{-1}\sum_{D_i=d} I(\hat{e}_i(\tau)>0)=(1-\tau)+O_p(n_d^{-1})=(1-\tau)+o_p((n_1+n_0)^{-1/2}).$$

By using Lemma 4.6 of He and Shao (1996), we will show (2.6). We introduce some notations first. Let $\{x_i, i\ge 1\}$ be independent random variables from probability distributions $F_{i,\theta}, i=1,\cdots,n$, with a common unknown parameter $\theta\in\Theta$, an open subset of $R^m, m\ge 1$. Consider a score function $\psi(x_i,\theta)$ with $\lambda_i(\theta)=E\psi(x_i,\theta)$ and $\Lambda_n(\theta)=\sum_{i=1}^n E\psi(x_i,\theta)$, and the M-estimator $\hat{\theta}_n$ of θ_0 that satisfies

$$\sum_{i=1}^n \psi(x_i,\hat{\theta}_n)=o(\delta_n),$$

where δ_n is a sequence of positive numbers. Let $u(x,\theta,d)=\sup_{|\nu-\theta|\le d}|\psi(x,\nu)-\psi(x,\theta)|$, where $|\cdot|$ is taken to be the sup norm defined as $|\theta|=\max(|\theta_1|,\cdots,|\theta_m|)$, and $Z_n(\nu,\theta)=|\sum_{i=1}^n(\psi(x_i,\nu)-\psi(x_i,\theta)-\lambda_i(\nu)+\lambda_i(\theta))|$. The conditions used in Lemma 4.6 of He and Shao (1996) are

(B1) For each fixed $\theta\in\Theta$, $\psi(x,\theta)$ is Borel measurable.

(B2) There exists $\theta_0\in\Theta$ such that $\Lambda_n(\theta_0)=0$ and $|\hat{\theta}_n-\theta_0|\to 0$ almost surely as $n\to\infty$.

(B3) There exist $r>0, d_0>0$ and a sequence of positive numbers $\{a_i, i\ge 1\}$ such that $Eu^2(x_i,\theta,d)\le a_i^2 d^r$ for $|\theta-\theta_0|\le d_0$ and $d\le d_0$.

(B4) $A_{2n}=O(A_n)$, where $A_n=\sum_{i=1}^n a_i^2$.

(B5') For some decreasing sequence of positive numbers d_n such that $d_n = O(d_{2n}) = o(1)$, $\max_{1\le i\le n} u(x_i, \theta_0, d_n) = O(A_n^{1/2} d_n^{r/2} (\log n)^{-2})$ a.s.

Lemma 4.6 of He and Shao (1996)

Assume that (B1), (B3) and (B5') are satisfied. Then we have

$$\limsup_{n\to\infty} \sup_{|\nu-\theta_0|\le d_n} \frac{Z_n(\nu,\theta_0)}{(A_n d_n^r + 1)^{1/2} (\log\log(n+A_n))^{1/2}} \le C \quad a.s.,$$

for some constant $C < \infty$.

Let $\theta_0 = 0, \psi(e_i(\tau), \theta) = e_i(\tau)\{I(e_i(\tau) > x_i^T\theta) - I(e_i(\tau) > 0)\}$, then

$$\begin{aligned}
\lambda_i(\theta) &= E\psi(e_i(\tau),\theta) = E\big(e_i(\tau)\{I(e_i(\tau) > x_i^T\theta) - I(e_i(\tau) > 0)\}\big),\\
Z_{n_d}(\nu,\theta_0) &= \big|\sum_{D_i=d} \{\psi(e_i(\tau),\nu) - \psi(e_i(\tau),\theta_0) - \lambda_i(\nu) + \lambda_i(\theta_0)\}\big|\\
&= \big|\sum_{D_i=d} e_i\{I(e_i(\tau) > x_i^T\nu) - I(e_i(\tau) > 0)\}\\
&\quad + \sum_{D_i=d} E\big(e_i(\tau)\{I(e_i(\tau) > 0) - I(e_i(\tau) > x_i^T\nu)\}\big)\big|.
\end{aligned}$$

First, we have

$$\begin{aligned}
&n_d^{-1} \sum_{D_i=d} E\big(e_i(\tau)\{I(e_i(\tau) > 0) - I(e_i(\tau) > x_i^T\nu)\}\big) = n_d^{-1} \sum_{D_i=d} \int_0^{x_i^T\nu} e_i(\tau) f(e_i(\tau)) de_i(\tau)\\
&= n_d^{-1} \sum_{D_i=d} x_i^T\nu\xi_i f(\xi_i) \le n_d^{-1} \sum_{D_i=d} (x_i^T\nu)^2 f(\xi_i) = \nu^T\big(\sum_{D_i=d} f(\xi_i) x_i x_i^T / n_d\big)\nu\\
&= O(\| \nu \|_2^2),
\end{aligned}$$

where ξ_i is between 0 and $x_i^T\nu$. Therefore,

$$n_d^{-1} Z_{n_d}(\nu,\theta_0) = \big|n_d^{-1} \sum_{D_i=d} e_i(\tau)\{I(e_i(\tau) > x_i^T\nu) - I(e_i(\tau) > 0)\}\big| + O(\| \nu \|_2^2).$$

Conditions (B1), (B3) and (B5') are checked as follows,

(B1) For each fixed θ, $\psi(e_i(\tau),\theta) = e_i(\tau)\{I(e_i(\tau) > x_i^T\theta) - I(e_i(\tau) > 0)\}$ is Borel measurable.

(B3) $u(e_i(\tau),\theta,d) = \sup_{|\nu-\theta|\le d} |e_i(\tau)\{I(e_i(\tau) > x_i^T\nu) - I(e_i(\tau) > x_i^T\theta)\}| = \sup_{|\nu-\theta|\le d} |e_i(\tau) I(x_i^T\nu <$

$e_i(\tau) < x_i^T\theta)| = |e_i(\tau)I(x_i^T\nu^* < e_i(\tau) < x_i^T\theta)|$, where $\nu^* = \theta - d(1, 1, \mathrm{sgn}(C_i^T))^T$. Therefore,

$$Eu^2(e_i(\tau), \theta, d) \quad = \quad \int_{x_i^T\nu^*}^{x_i^T\theta} e_i^2(\tau)f(e_i(\tau))de_i(\tau) \le M(x_i^T\theta)^2 \parallel x_i \parallel_1 d \le Md_0^2 \parallel x_i \parallel_1^3 d,$$

where $|\theta| \le d_0$, and $\parallel x_i \parallel_1 = 1 + D_i + \parallel C_i \parallel_1$. Condition (B3) holds if we take $r = 1$, and $a_i^2 = Md_0^2 \parallel x_i \parallel_1^3$.

(B5') Let $d_{n_d} = n_d^{-1/2}\log n_d$, we have

$$\begin{aligned}\frac{\max\limits_{1\le i\le n_d} u(e_i(\tau), \theta_0, d_{n_d})}{A_{n_d}^{1/2}d_{n_d}^{1/2}(\log n_d)^{-2}} &= \frac{\max\limits_{1\le i\le n_d} |e_i(\tau)I(x_i^T\nu^* < e_i(\tau) < 0)|}{\{Md_0^2\sum_{D_i=d} \parallel x_i \parallel_1^3\}^{1/2}d_{n_d}^{1/2}(\log n_d)^{-2}} \\ &\le \frac{d_{n_d}\max_{1\le i\le n_d} \parallel x_i \parallel_1}{M^{1/2}d_0(\sum_{D_i=d} \parallel x_i \parallel_1^3)^{1/2}d_{n_d}^{1/2}(\log n_d)^{-2}} \\ &= M^{-1/2}d_0^{-1}\frac{\max_{1\le i\le n_d} \parallel x_i \parallel_1 n_d^{-1/2}}{(\sum_{D_i=d} \parallel x_i \parallel_1^3 /n_d)^{1/2}}d_{n_d}^{1/2}(\log n_d)^2 \to 0 \text{ a.s.} \\ &\quad \text{as } n_d \to \infty,\end{aligned}$$

where $\max_{1\le i\le n_d} \parallel x_i \parallel_1 \le 2 + \max_{1\le i\le n_d} \parallel C_i \parallel_1 = O(n_d^{1/2})$ according to Lemma 11.2 in Owen (2001, p.218), $\sum_{D_i=d} \parallel x_i \parallel_1^3 /n_d$ is bounded away from 0, and $d_{n_d}^{1/2}(\log n_d)^2 = o(1)$.

Lemma 11.2 of Owen (2001)

Let Y_i be independent random variables with a common distribution and $E(Y_i^2) < \infty$. Let $Z_n = \max_{1\le i\le n}|Y_i|$. Then $Z_n = o(n^{1/2})$.

Because (B1), (B3) and (B5') are satisfied, according to Lemma 4.6 of He and Shao (1996), we have

$$\limsup_{n_d\to\infty} \sup_{|\nu-\theta_0|\le d_{n_d}} \frac{n_d^{-1}Z_{n_d}(\nu, \theta_0)}{n_d^{-1}(A_{n_d}d_{n_d}^r + 1)^{1/2}(\log\log(n_d + A_{n_d}))^{1/2}} \le C \quad a.s.$$

The denominator is

$$\begin{aligned}&n_d^{-1}(A_{n_d}d_{n_d}^r + 1)^{1/2}(\log\log(n_d + A_{n_d}))^{1/2} \\ &= n_d^{-1}\big(Md_0^2\sum_{D_i=d} \parallel x_i \parallel_1^3 n_d^{-1}(n_d d_{n_d}) + 1\big)^{1/2}\big(\log\log(n_d + Md_0^2\sum_{D_i=d} \parallel x_i \parallel_1^3)\big)^{1/2} \\ &= O(n_d^{-1}(n_d n_d^{-1/2}\log n_d)^{1/2}(\log\log n_d)^{1/2}) = o((n_1 + n_0)^{-1/2}),\end{aligned}$$

where $\sum_{D_i=d} \parallel x_i \parallel_1^3 n_d^{-1} = O(1)$.

Therefore, the numerator is

$$n_d^{-1} Z_{n_d}(\nu, \theta_0) = o_p((n_1 + n_0)^{-1/2}), \text{uniformly in } \{\nu : |\nu - \theta_0| \leq d_{n_d}\}.$$

Take $\nu = \hat{\beta}(\tau) - \beta(\tau)$, we have

$$n_d^{-1} \big[\sum_{D_i=d} e_i(\tau)\{I(\hat{e}_i(\tau) > 0) - I(e_i(\tau) > 0)\} \big] + O(\parallel \hat{\beta}(\tau) - \beta(\tau) \parallel_2^2) = o_p((n_1 + n_0)^{-1/2}),$$

where $O(\parallel \hat{\beta}(\tau) - \beta(\tau) \parallel_2^2) = O((n_1 + n_0)^{-1}) = o_p((n_1 + n_0)^{-1/2})$.

Therefore, we have

$$n_d^{-1} \big[\sum_{D_i=d} e_i(\tau)\{I(\hat{e}_i(\tau) > 0) - I(e_i(\tau) > 0)\} \big] = o_p((n_1 + n_0)^{-1/2}).$$

Theorem 2.1.1. *Suppose that* $\lim_{n_1,\ n_0 \to \infty} (n_1 + n_0)^{-1} U_f$ *exists,* $E \parallel C_i \parallel_1^3 < \infty$*, and* f_i *are uniformly bounded away from 0 and infinity. Under the null hypothesis that* $F_{Z|C,D=0} = F_{Z|C,D=1}$*, we have*

$$T_\tau^{COVES}(n_1,\ n_0)/s_{n_1,n_0} \to N(0,1), \qquad \textit{as } n_1,\ n_0 \to \infty.$$

Proof of Theorem 2.1.1 According to (2.3), Lemma 2.1.1, and $\delta(\tau) = 0$ under H_0, we have

$$\begin{aligned} T_\tau(n_1, n_0) &= \sum_i \Big[\big(I(D_i = 1) - I(D_i = 0)\big)\{(1-\tau)n_d\}^{-1} e_i(\tau) I(e_i(\tau) > 0) - (\bar{C}_\tau(1) - \bar{C}_\tau(0))^T U_f^{-1} C_i^* \psi_\tau(e_i(\tau)) \Big] \\ &\quad + o_p((n_1 + n_0)^{-1/2}) \\ &= T^*(n_1, n_0) + o_p((n_1 + n_0)^{-1/2}), \end{aligned}$$

where

$$T_\tau^*(n_1, n_0) = \sum_i \Big[\big(I(D_i = 1) - I(D_i = 0)\big)\{(1-\tau)n_d\}^{-1} e_i(\tau) I(e_i(\tau) > 0) - (\bar{C}_\tau(1) - \bar{C}_\tau(0))^T U_f^{-1} C_i^* \psi_\tau(e_i(\tau)) \Big].$$

We note that

$$\begin{aligned} E(T^*_\tau(n_1,n_0)) &= (1-\tau)^{-1}\Big\{n_1^{-1}\sum_{D_i=1} E\big(e_i(\tau)I(e_i(\tau)>0)\big) - n_0^{-1}\sum_{D_i=0} E\big(e_i(\tau)I(e_i(\tau)>0)\big)\Big\} \\ &= (1-\tau)^{-1}E\big(e(\tau)I(e(\tau)>0)\big)(1-1)=0. \end{aligned}$$

Further, we have

$$\begin{aligned} \mathrm{var}(T^*_\tau(n_1,n_0)) &= (1-\tau)^{-2}\mathrm{Var}\big(e(\tau)I(e(\tau)>0)\big)(n_1^{-1}+n_0^{-1}) \qquad (2.7) \\ &\quad +\tau(1-\tau)(\bar{C}_\tau(1)-\bar{C}_\tau(0))^T U_f^{-1}\{\sum_i C_i^* C_i^{*T}\}U_f^{-1}(\bar{C}_\tau(1)-\bar{C}_\tau(0)), \qquad (2.8) \end{aligned}$$

which can be estimated by $s^2_{n_1,n_0}$.

By the Central Limit Theorem, $T^*_\tau(n_1,\ n_0)$ is asymptotically normal with mean 0 and variance (2.7). Using the results from Lemma 2.1.1 and $T^{COVES}_\tau(n_1,\ n_0) = T^*_\tau(n_1,\ n_0) + o_p((n_1+n_0)^{-1/2})$, the proof is complete. $\diamond$

To use the asymptotic normality for testing the null hypothesis of no treatment effects, we need a consistent estimate of U_f. If $e_i(\tau)$ follows a common distribution, then a kernel density estimate can be used to estimate the common density at 0. If the conditional densities vary with C_i, it is not possible to estimate each f_i consistently, but U_f, a linear cobmination of the f_i's can still be consistently estimated; see He et al. (2002) and Koenker (2005) for more details. For the simulations in this chapter, the proposed test is carried out using a kernel density estimate, *density*, in R.

$$\hat{f}_i = \frac{1}{(n_1+n_0)h_{n_1+n_0}}\sum_i K(\frac{0-\hat{e}_i(\tau)}{h_{n_1+n_0}}),$$

where K is a kernel function, and $h_{n_1+n_0}$ is a bandwidth, also called a smoothing parameter. In the simulation, we used a rule of thumb bandwidth of a Gaussian kernel density provided by Silverman (1986), that is

$$h_{n_1+n_0} = 0.9A(n_1+n_0)^{-1/5},$$

where $A = \min\{\text{standard deviation}, \text{interquartile range}/1.34\}$.

2.2 Simulations

Suppose that Z is the response variable under two treatments according to distributions F_1 and F_0. We consider the problem of testing the null hypothesis that $F_{Z|C,D=0} = F_{Z|C,D=1}$ versus the two-sided alternative hypothesis when some covariates, C, may affect outcome measurements. For simplicity, we consider C as a univariate covariate in the simulation. Five tests are considered to detect the treatment effect, where the difference may lie in the upper tail of the distributions. The $0.5^{\text{th}}/0.75^{\text{th}}$ COVES test based on the difference in the covariate-adjusted expected loss above the $0.5^{\text{th}}/0.75^{\text{th}}$ quantile function, the $0.5^{\text{th}}/0.75^{\text{th}}$ ES test based on the difference in the expected loss above the $0.5^{\text{th}}/0.75^{\text{th}}$ quantile without adjusting for covariates, and the t test based on the least squares regression.

2.2.1 Normal Models

We consider data generated from

$$Z_i = 5 + \gamma C_i + \{1 + \eta I(e_i > 0) I(D_i = 0)\} e_i$$

where $e_i \sim N(0, 1)$, and η is either 0 (under the null hypothesis) or 1.35 (under the alternative hypothesis). The coefficient γ and the distribution for the covariate C_i will be specified later. Clearly, the control group ($d = 0$) has a heavier right tail. As we discuss in Subsection 1.4.1, when $\eta = 1.35$, the error variance of the control group ($d = 0$) is about triple that of the treatment group ($d = 1$) under this model. We will consider four scenarios for the effects of the covariate in the analysis.

- *Scenario 1, no covariate effect:* we take C_i from $N(2.5,\ 0.5^2)$, with $\gamma = 0$.
- *Scenario 2, a common covariate effect:* we take C_i from $N(2.5,\ 0.5^2)$, with $\gamma = 1$.
- *Scenario 3, a covariate distribution that varies with treatment groups:* we take C_i from $N(2.5,\ 0.5)$ for $d = 0$, but from $N(3.0, 0.5^2)$ for $d = 1$, with $\gamma = 1$.
- *Scenario 4, a covariate distribution that has a scale change across treatment groups:* we take C_i from $N(2.5,\ 0.5^2)$ for $d = 0$, but from $N(2.5, 1.0)$ for $d = 1$, with $\gamma = 1$.

Type I Error and Histogram of p-values

A total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 50$ are generated for Z under the null hypothesis (with $\eta = 0$). From Table 2.1, Figure 2.1, and Figure 2.2, we find that the p-values for the ES test in Scenario 3 and Scenario 4 deviates a lot from the uniform distribution, which means that the ES test is not appropriate. It is clear that when the covariate between two groups differs a lot in location or scale, we can not ignore the covariate effect when we analyze the data.

Table 2.1: Type I error at sample size $(n_1,\ n_0) = (50, 50)$ and nominal level 5% for for i.i.d errors under normal covariates. The values in the table are in percentage.

Scenario	0.5th COVES test	0.75th COVES test	0.5th ES test	0.75th ES test	t test
1	5.70	6.79	5.41	5.15	5.17
2	5.73	6.76	4.97	4.68	5.17
3	5.34	6.44	**46.78**	**32.89**	5.03
4	5.96	7.02	**11.41**	**14.79**	5.12

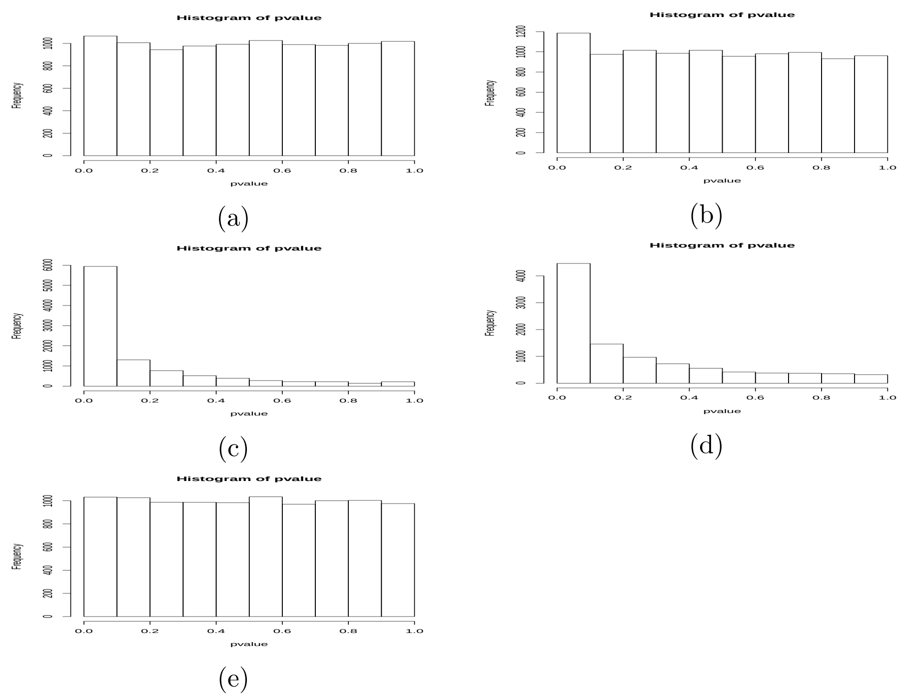


Figure 2.1: Histograms of p-values in Scenario 3, a covariate distribution that varies with treatment groups for i.i.d errors. (a) 0.5th COVES test. (b) 0.75th COVES test. (c) 0.5th ES test. (d) 0.75th ES test. (e) t test.

Power Function

We will compare power functions for five tests, the $0.5^{\text{th}}/0.75^{\text{th}}$ COVES tests, the $0.5^{\text{th}}/0.75^{\text{th}}$ ES tests, and the t test, and then provide the sample sizes needed for each test to reach power 0.9 at a given alternative. The grid is 10 for sample size, and we use smoothing splines to obtain power curves. In Scenarios 3 and 4, the adjustment of the covariate is important. Because the ES test is not valid, we will not include the ES test in these two scenarios. The results clearly show the superiority of the COVES test. Consider Scenarios 1 and 2, when there is a covariate effect

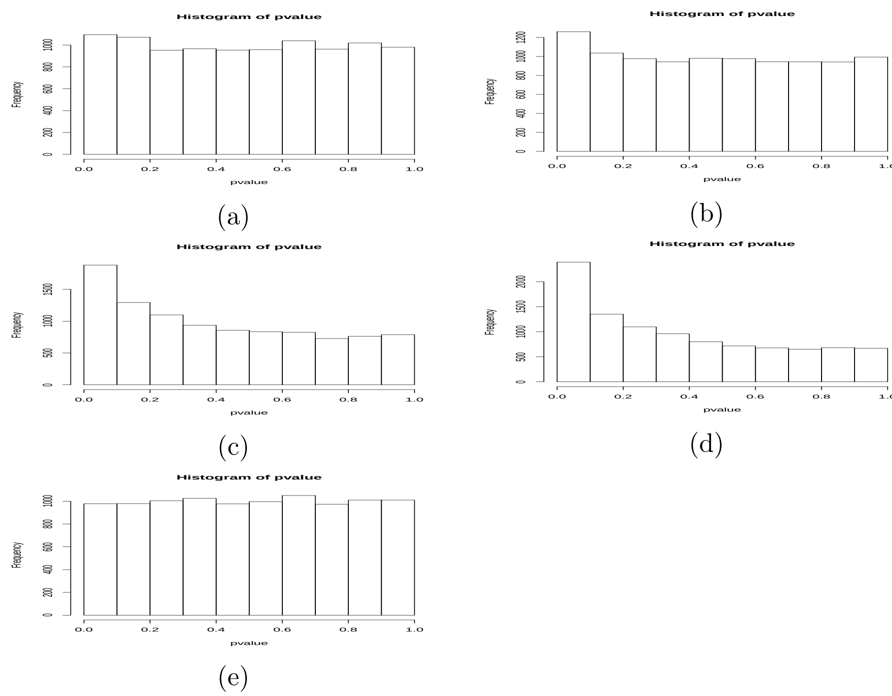


Figure 2.2: Histograms of p-values in Scenario 4, a covariate distribution that has a scale change across treatment groups for i.i.d errors. (a) 0.5th COVES test. (b) 0.75th COVES test. (c) 0.5th ES test. (d) 0.75th ES test. (e) t test.

(Scenario 2), the COVES test is more powerful. So, it's more effective to compare the ES difference after adjusting the covariates. Even when there is no covariate effect (Scenario 1), the COVES test is nearly as good as the ES test. The t test required a sample size about twice as large.

Table 2.2: Comparisons of five tests. (a) Type I error at sample size $(n_1,\ n_0) = (50, 50)$ and nominal level 5% for for i.i.d errors. The values in the table are in percentage. (b) Sample size $(n_1,\ n_0)$ needed to reach power 0.9 at the alternative. In each scenario, the first row uses $n_1 = n_0$, and the second row uses $n_1 = 2n_0$.

	0.5th COVES test		0.75th COVES test		0.5th ES test		0.75th ES test		t test	
Scenario	(a)	(b)	(a)	(b)	(a)	(b)				
1	5.70	(75, 75)	6.79	(51, 51)	5.41	(74, 74)	5.15	(50, 50)	5.17	(140, 140)
		(140, 70)		(92, 46)		(136, 68)		(94, 47))		(202, 101)
2	5.73	(76, 76)	6.76	(51, 51)	4.97	(83, 83)	4.68	(59, 59)	5.17	(140, 140)
		(138, 69)		(92, 46)		(148, 74)		(110, 55)		(202, 101)
3	5.34	(88, 88)	6.44	(59, 59)	**46.78**		**32.89**		5.03	(177, 177)
		(152, 76)		(100, 50)						(240, 120)
4	5.96	(77, 77)	7.02	(51, 51)	**11.41**		**14.79**		5.12	(141, 141)
		(140, 70)		(96, 48)						(202, 101)

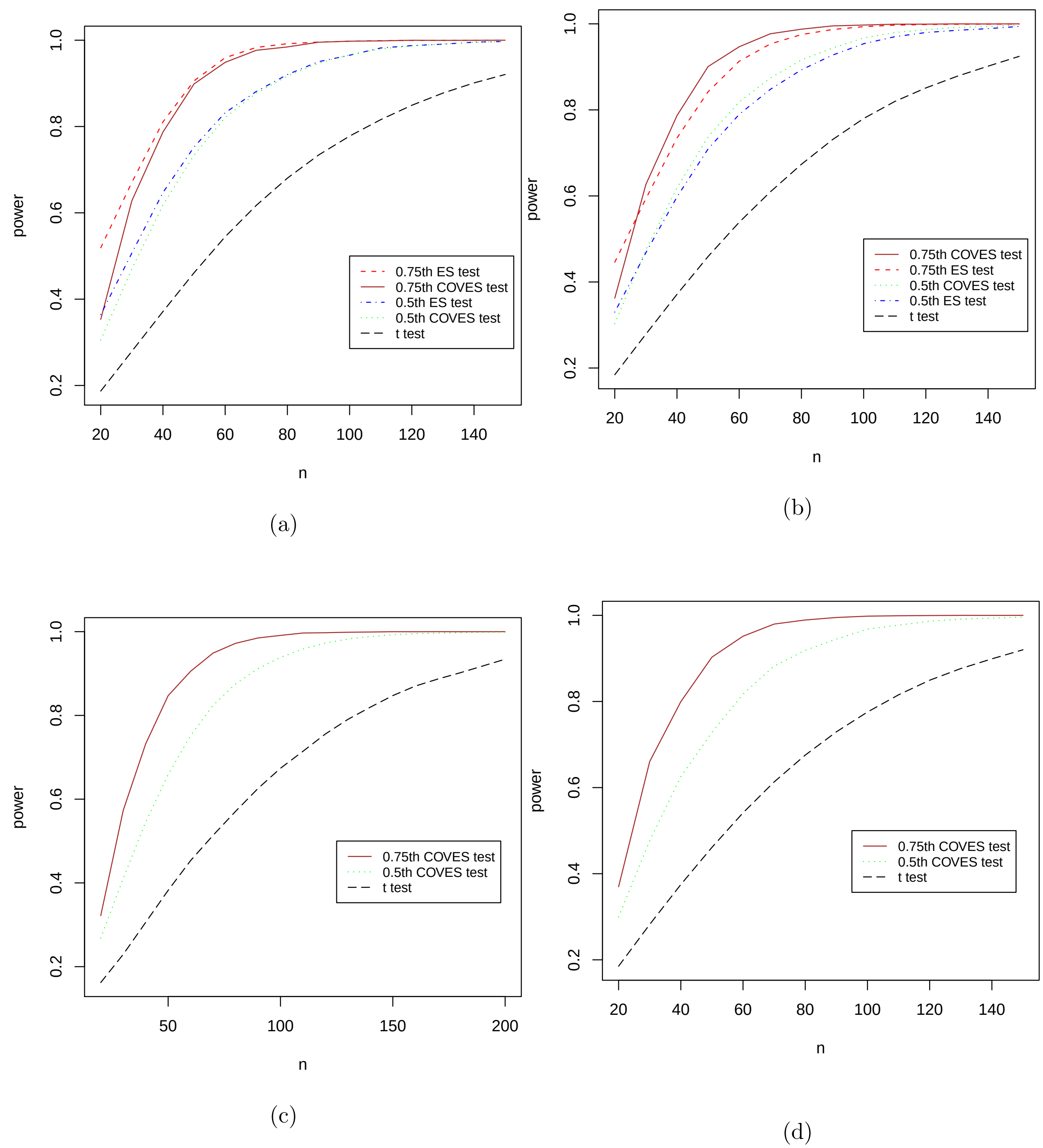


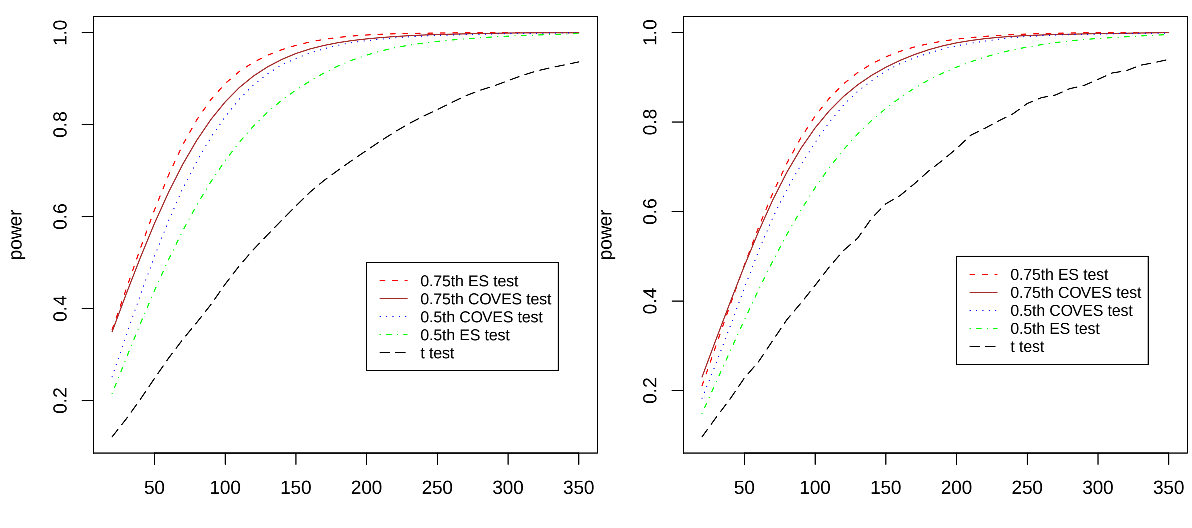
Figure 2.3: Statistical powers of five tests as functions of sample size $n_1 = n_0 = n$ for i.i.d errors. (a) Scenario 1. (b) Scenario 2. (c) Scenario 3. (d) Scenario 4.

2.2.2 Targeted Study on TSS

We use the same data discussed in Subsection 1.4.2, but use the baseline TSS as the covariate in the analysis. For the control group, the baseline TSS is generated in such a way that the quantiles of the baseline TSS in the two groups are the same and the outcome measurement is simulated according to the same two scenarios discussed in Subsection 1.4.2.

Table 2.3 reports the type I errors at a total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 100$. It also reports the sample sizes needed to reach power of 0.90 in each scenario under two design conditions: $n_1 = n_0$ and $n_1 = 2n_0$, respectively. The results clearly show the baseline TSS does not play a significant role, so the statistical power for detecting the treatment effect has no gain (0.75th quantile) or just gain a little bit (0.5th quantile) by adjusting the covariate in the analysis. However, the COVES test is clearly outperforming the t test based on the least squares regression, and the latter would require a trial that is more than double in size. In both scenarios, the 0.75th COVES/ES tests outperform the 0.5th COVES/ES test because the difference between two groups lies in the upper tail of the 0.65th and 0.75th quantiles, respectively.

The power functions for these tests in each scenario are shown in Figure 2.4 with sample sizes up to $n_1 = n_0 = 350$. The grid is 10 for sample size, and we use smoothing splines to obtain power curves. The t test required more than double sample size to reach power 0.9.



(a) (b)

Figure 2.4: Statistical powers of five tests in reference with the TSS data with agreement up to the (a) 0.65th quantile and (b) 0.75th quantile, as functions of sample size $n_1 = n_0 = n$.

Table 2.3: Comparisons of five tests. (a) Type I error at sample size $(n_1,\ n_0) = (100, 100)$ and nominal level 5%. The values in the table are in percentage. (b) Sample size (n_1, n_0) needed to reach power 0.9. In each scenario, the first row uses $n_1 = n_0$, and the second row uses $n_1 = 2n_0$.

	0.5th COVES test		0.75th COVES test		0.5th ES test		0.75th ES test		t test	
Scenario	(a)	(b)	(a)	(b)	(a)	(b)				
1	4.85	(129, 129)	6.78	(120, 120)	5.41	(170, 170)	4.96	(110, 110)	4.99	(306, 306)
		(202, 101)		(172, 86)		(276, 138)		(158, 79)		(450, 225)
2	4.85	(145, 145)	6.78	(122, 122)	5.41	(188, 188)	4.96	(130, 130)	4.99	(305, 305)
		(250, 125)		(246, 123)		(322, 161)		(230, 115)		(442, 221)

2.3 Conclusion

We have proposed a new test to detect treatment effects where the difference between two treatment groups may lie in the upper tail of the distributions. The χ^2 test mentioned in Section 1.4 is difficult to implement if we need to adjust for covariates. Besides, there has been no agreeable cutoff point for dichotomization. The t test based on the least squares regression focuses on the conditional mean, which is not powerful for detecting the difference appearing in the tail. In contrast, we take the covariate-adjusted outcomes above the τ-th quantile regression line to construct the COVES test statistic, which compares the covariate-adjused expected shortfalls between two groups. Empirical studies illustrate that the proposed test has a clear advantage in reducing sample sizes in clinical trials.

Chapter 3

Generalized Covariate-adjusted Expected Shortfall Test

The COVES test adjusts for covariates, and is shown to be valid for i.i.d (independent and identically distributed) error models. We show that the COVES test is valid when the covariates have the same means across treatments in non-iid errors. However, in a non-randomized design and non-i.i.d error model, the covariates may not have the same means across treatments. The COVES test is not valid for this scenario and we propose the Generalized Covariate-adjusted Expected Shortfall (q.COVES) to generalizes the COVES test to more general models.

Given data (Z_i, D_i, C_i) with $D_i = 1$ for $i = 1, \cdots, n_1$ and $D_i = 0$ for $i = n_1 + 1, \cdots, n_1 + n_0$. Consider a quantile model

$$\begin{aligned} Z_i &= \alpha(u_i) + D_i\delta(u_i) + C_i^T\gamma(u_i) = x_i^T\beta(u_i) \\ &= \alpha(\tau) + D_i\delta(\tau) + C_i^T\gamma(\tau) + e_i(\tau) = x_i^T\beta(\tau) + e_i(\tau), \end{aligned} \tag{3.1}$$

where $x_i = (1, D_i, C_i^T)^T$, $\beta(u) = (\alpha(u), \delta(u), \gamma^T(u))^T$, $e_i(\tau)$ are independent but not identically distributed, with

$$\begin{aligned} &u_i \sim U(0,1), \\ &\alpha(u_i) = \Phi^{-1}(u_i), \\ &e_i(\tau) = (\alpha(u_i) - \alpha(\tau)) + D_i(\delta(u_i) - \delta(\tau)) + C_i^T(\gamma(u_i) - \gamma(\tau)) = x_i^T(\beta(u_i) - \beta(\tau)), \end{aligned}$$

where Φ is the cumulative distribution function of the standard normal distribution. Here, u_i is the quantile level in the conditional distribution of Z_i given (D_i, C_i), and $\{e_i(\tau) > 0\}$ is equivalent to $\{u_i > \tau\}$ given (D_i, C_i). The distribution for the covariates C_i will be specified later. We consider the hypothesis

$$
\begin{aligned}
H_0 &: \delta(u) = 0 \text{ for all } u \in (0,1) \\
\text{v.s. } H_1 &: \delta(u) = \begin{cases} a(u-1/2) & \text{, if } u > 1/2 \\ 0 & \text{, otherwise.} \end{cases}
\end{aligned}
$$

for some $a > 0$. Let $\hat{\beta}(\tau) = (\hat{\alpha}(\tau), \hat{\delta}(\tau), \hat{\gamma}^T(\tau))$ be the regression quantile coefficients at the τ-th quantile level, S_d be the number of Z_i in group d that are above the τ-th regression quantile, $\hat{e}_i(\tau)$ be the residuals from the τ-th conditional quantile given $(D_i,\ C_i)$. Then

$$S_d = \sum_{D_i=d} I(\hat{e}_i(\tau) > 0).$$

Let

$$
\begin{aligned}
Y_i &= Z_i - \hat{\alpha}(\tau) - C_i^T \hat{\gamma}(\tau) = (\alpha(u_i) - \hat{\alpha}(\tau)) + D_i \delta(u_i) + C_i^T(\gamma(u_i) - \hat{\gamma}(\tau)) \\
&= \tilde{x}_i^T(\beta(u_i) - \hat{\beta}(\tau)) + D_i \delta(u_i)
\end{aligned}
$$

be the covariate-adjusted outcome, where $\tilde{x}_i = (1, 0, C_i^T)^T$. We assign each covariate-adjusted outcome, Y_i, a weight $q_{d,i}$ to develop a new COVES test statistic such that the mean difference in C_i will be accounted for. Here, the generalized COVES (abbreviated as q.COVES) test statistic is

$$T_\tau^{q.COVES}(n_1, n_0) = \sum_{D_i=1} w_{d,i} q_{d,i} Y_i - \sum_{D_i=0} w_{d,i} q_{d,i} Y_i, \tag{3.2}$$

where

$$w_{d,i} = S_d^{-1} I(\hat{e}_i(\tau) > 0).$$

According to the Bahadur representation of the quantile estimator (Koenker, 2005), we have

$$\hat{\beta}(\tau) - \beta(\tau) = H_{n_1+n_0}^{-1}(n_1+n_0)^{-1} \sum_i x_i \psi_\tau(e_i(\tau)) + (n_1+n_0)^{-1/2} R_{n_1+n_0},$$

where $\hat{\beta}(\tau) = (\hat{\alpha}(\tau), \hat{\delta}(\tau), \hat{\gamma}^T(\tau))^T$, $\beta(\tau) = (\alpha(\tau), \delta(\tau), \gamma^T(\tau))^T$, $x_i = (1, D_i, C_i^T)^T$, $\psi_\tau(e_i(\tau)) = \tau - I(e_i(\tau) < 0)$, f_i is the conditional density function of $e_i(\tau)$ given (D_i, C_i) evaluated at 0, the Hessian matrix $H_{n_1+n_0} = \lim\limits_{n_1,\ n_0 \to \infty} (n_1+n_0)^{-1} \sum_i f_i x_i x_i^T$, and the remainder term $R_{n_1+n_0} = o_p(1)$.

The matrix $H_{n_1+n_0}$ can be consistently estimated by the sparsity estimation method (Hendricks and Koenker, 1992), which uses

$$\hat{f}_i = 2h_{n_1+n_0}/\{x_i^T(\hat{\beta}(\tau + h_{n_1+n_0}) - \hat{\beta}(\tau - h_{n_1+n_0}))\},$$

where the bandwidth

$$h_{n_1+n_0} = (n_1+n_0)^{-1/5}\left[\frac{4.5\phi^4(\Phi^{-1}(\tau))}{(2\Phi^{-1}(\tau)^2+1)^2}\right]^{1/5}.$$

Let

$$H^{-1} = (\sum_i f_i x_i x_i^T)^{-1}.$$

We have,

$$\hat{\beta}(\tau) - \beta(\tau) \quad = \quad H^{-1}\sum x_i\psi_\tau(e_i(\tau)) + o_p((n_1+n_0)^{-1/2}). \qquad (3.3)$$

Theorem 3.0.1. *Suppose that* $\lim_{n_1,\ n_0\to\infty}(n_1+n_0)^{-1}H$ *exists. Under the null hypothesis that* $F_{Z|C,D=0} = F_{Z|C,D=1}$*, we have*

$$T_\tau^{q.COVES}(n_1,\ n_0)/s_{n_1,n_0} \to N(0,1), \qquad as\ n_1,\ n_0 \to \infty,$$

where

$$\begin{aligned}
s^2_{n_1,n_0} \quad &= \quad (1-\tau)^{-2}\sum_i n_d^{-2}q_{d,i}^2\tilde{x}_i^T V_\beta \tilde{x}_i + \tau(1-\tau)(\overline{Xq}_\tau(1) - \overline{Xq}_\tau(0))^T H^{-1}\{\sum_i x_i x_i^T\}H^{-1}(\overline{Xq}_\tau(1) - \overline{Xq}_\tau(0)) \\
&\quad -2\tau(1-\tau)^{-1}\sum_i\big\{(I(D_i=1) - I(D_i=0))n_d^{-1}q_{d,i}\tilde{x}_i^T E_\beta x_i^T H^{-1}(\overline{Xq}_\tau(1) - \overline{Xq}_\tau(0))\big\}, \\
\tilde{x}_i \quad &= \quad (1,\ 0, C_i^T)^T, \\
\overline{Xq}_\tau(d) \quad &= \quad S_d^{-1}\sum_{D_i=d} q_{d,i}\tilde{x}_i I(\hat{e}_i(\tau) > 0), \\
E_\beta \quad &= \quad (n_1+n_0)^{-1}\sum_i\{(\hat{\beta}(u_i) - \hat{\beta}(\tau))I(u_i > \tau)\},\ \ and \\
V_\beta \quad &= \quad (n_1+n_0)^{-1}\sum_i\{(\hat{\beta}(u_i) - \hat{\beta}(\tau))(\hat{\beta}(u_i) - \hat{\beta}(\tau))^T I(u_i > \tau)\} - E_\beta^2.
\end{aligned}$$

Proof of Theorem 3.0.1 According to (3.2), (3.3), and $\delta(u) = 0$ under H_0, we have

$$\begin{aligned}
T_\tau^{q.COVES}(n_1, n_0) &= \sum_i \big\{(I(D_i = 1) - I(D_i = 0))S_d^{-1} q_{d,i}\tilde{x}_i^T(\beta(u_i) - \hat{\beta}(\tau))I(\hat{e}_i(\tau) > 0)\big\} \\
&= \sum_i \big[(I(D_i = 1) - I(D_i = 0))\{(1-\tau)n_d\}^{-1} q_{d,i}\tilde{x}_i^T(\beta(u_i) - \beta(\tau))I(u_i > \tau) \\
&\quad -(\overline{Xq}_\tau(1) - \overline{Xq}_\tau(0))^T H^{-1} x_i \psi_\tau(e_i(\tau))\big] + o_p((n_1 + n_0)^{-1/2}) \\
&= T^*(n_1, n_0) + o_p((n_1 + n_0)^{-1/2}),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{x}_i &= (1,\ 0, C_i^T)^T,\ \overline{Xq}_\tau(d) = S_d^{-1} \sum_{D_i = d} q_{d,i}\tilde{x}_i I(\hat{e}_i(\tau) > 0),\ \text{and} \\
T^*(n_1, n_0) &= \sum_i \big[(I(D_i = 1) - I(D_i = 0))\{(1-\tau)n_d\}^{-1} q_{d,i}\tilde{x}_i^T(\beta(u_i) - \beta(\tau))I(u_i > \tau) \\
&\quad -(\overline{Xq}_\tau(1) - \overline{Xq}_\tau(0))^T H^{-1} x_i \psi_\tau(e_i(\tau))\big].
\end{aligned}$$

We note that

$$E(T^*(n_1, n_0)) = (1-\tau)^{-1}\big(n_1^{-1} \sum_{D_i=1} q_{d,i}\tilde{x}_i - n_0^{-1} \sum_{D_i=0} q_{d,i}\tilde{x}_i\big)^T E\big((\beta(u) - \beta(\tau))I(u > \tau)\big).$$

If covariates have the same means across treatments, then

$$n_1^{-1} \sum_{D_i=1} \tilde{x}_i = n_0^{-1} \sum_{D_i=0} \tilde{x}_i.$$

We have $E(T^*(n_1, n_0)) = 0$ with the choice of $q_{d,i} = 1$ for all i. In this scenario, the q.COVES test reduces to the COVES test. On the other hand, if any covariate has different means across treatments, then

$$n_1^{-1} \sum_{D_i=1} \tilde{x}_i \neq n_0^{-1} \sum_{D_i=0} \tilde{x}_i.$$

We have $E(T^*(n_1, n_0)) \neq 0$ with the choice of $q_{d,i} = 1$ for all i. To ensure $E(T^*(n_1, n_0)) = 0$, we need to choose weights to satisfy

$$n_1^{-1} \sum_{D_i=1} q_{d,i}\tilde{x}_i = n_0^{-1} \sum_{D_i=0} q_{d,i}\tilde{x}_i. \tag{3.4}$$

The choice of $q_{d,i}$ will be discussed in Section 3.1.1 to 3.1.3. Further, we have

$$\begin{aligned}
Var(T^*(n_1,n_0)) &= (1-\tau)^{-2}\sum_i n_d^{-2}q_{d,i}^2\tilde{x}_i^T\mathrm{Var}\big((\beta(u)-\beta(\tau))I(u>\tau)\big)\tilde{x}_i \\
&\quad +\tau(1-\tau)(\overline{Xq}_\tau(1)-\overline{Xq}_\tau(0))^T H^{-1}\{\sum_i x_i x_i^T\}H^{-1}(\overline{Xq}_\tau(1)-\overline{Xq}_\tau(0)) \\
&\quad -2\tau(1-\tau)^{-1}\sum_i\big\{(I(D_i=1)-I(D_i=0))n_d^{-1}q_{d,i}\tilde{x}_i^T E\big((\beta(u)-\beta(\tau))I(u>\tau)\big)x_i^T H^{-1} \\
&\quad (\overline{Xq}_\tau(1)-\overline{Xq}_\tau(0))\big\},
\end{aligned}$$

which can be estimated by $s^2_{n_1,n_0}$. We estimate the quantile level for each observation, u_i, by finding the level so that the quantile regression fit under H_1 passes the observation. For the ith observation, $i=1,\cdots,n_1+n_0$, the corresponding u_i is its qunatile level. Finally, $E\big((\beta(u)-\beta(\tau))I(u>\tau)\big)$ and $\mathrm{Var}\big((\beta(u)-\beta(\tau))I(u>\tau)\big)$ can be estimated by E_β and V_β, respectively. By the Central Limit Theorem, and $T_\tau^{q.COVES}(n_1,\ n_0)=T_\tau^*(n_1,\ n_0)+o_p((n_1+n_0)^{-1/2})$, the proof is complete. $\diamond$

There are two other ways to calculate the variance of $T^*(n_1,n_0)$. Let

$$\begin{aligned}
A_{di} &= (I(D_i=1)-I(D_i=0))\{(1-\tau)n_d\}^{-1}q_{d,i}\tilde{x}_i^T(\hat{\beta}(u_i)-\hat{\beta}(\tau))I(u_i>\tau) \\
&\quad -(\overline{Xq}_\tau(1)-\overline{Xq}_\tau(0))^T H^{-1}x_i\psi_\tau(e_i(\tau)).
\end{aligned}$$

We have

$$\begin{aligned}
\hat{T}^*(n_1,n_0) &= \sum_i A_{di} \\
&= \sum_{D_i=1}A_{1i}+\sum_{D_i=0}A_{0i},
\end{aligned}$$

and $Var(T^*(n_1,n_0))$ can be estimated either by

$$\sum_i A_{di}^2-(n_1+n_0)^{-1}(\sum_i A_{di})^2.$$

or

$$\sum_{d=0}^{1}\{\sum_{D_i=d}A_{di}^2-n_d^{-1}(\sum_{D_i=1}A_{di})^2\}.$$

According to the simulation results, we find that the the variance estimates from these two methods are not very stable. Therefore, we use $s^2_{n_1,n_0}$ for the variance calculation.

3.1 Simulations

Suppose that Z is the response variable under two treatments with distributions F_1 and F_0, respectively. We consider the problem of testing the null hypothesis that $F_{Z|C,D=0} = F_{Z|C,D=1}$ versus the two-sided alternative hypothesis, where the covariates C may affect treatment outcomes. For simplicity, we consider C as a univariate covariate in the simulation. We provide three studies for i.i.d error and non-i.i.d error models. The first study is i.i.d errors with normal covariates. This study shows that the COVES test is still valid under non-randomized design if the errors are i.i.d. The other two studies involve non-i.i.d errors with discrete and normal covariates, respectivelty. These two studies show that the COVES test is not valid in non-randomized design for non-i.i.d errors. Koenker (2010) proposed rank tests designed for heterogeneous treatment effect models. The rank tests use the regression rankscores introduced by Gutenbrunner and Jureckova (1992) to tackle covariate effects. In our simulations, we will include the rank test for comparison.

We will compare the $0.5^{\text{th}}/0.75^{\text{th}}$ q.COVES tests based on the difference in the q-weighted covariate-adjusted expected loss above the $0.5^{\text{th}}/0.75^{\text{th}}$ quantile function, the $0.5^{\text{th}}/0.75^{\text{th}}$ COVES tests based on the difference in the covariate-adjusted expected loss above the $0.5^{\text{th}}/0.75^{\text{th}}$ quantile function, the $0.5^{\text{th}}/0.75^{\text{th}}$ rank (aka, trimmed Wilcoxon) tests based on the difference in the integrated rankscore above $0.5^{\text{th}}/0.75^{\text{th}}$ quantile, and the t test based on the least squares regression.

3.1.1 Simulation Studies for i.i.d Errors with Normal Covariates

Thus far in this chapter, we have proposed the q.COVES test that works in non-i.i.d error models. There are several ways to choose weights to satisfy (3.4). We will provide a way later in Section 3.1.1 to 3.1.3. In this section, we show that both q.COVES and COVES tests work in i.i.d error models.

We consider data generated from

$$Z_i = 5 + \gamma C_i + \{1 + \eta I(e_i > 0)I(D_i = 0)\}e_i$$

where $e_i \sim N(0, 1)$, and η is 0 under the null hypothesis. The coefficient γ and the distribution for the covariate C_i will be specified later. We consider four scenarios for the effects of the covariate in the analysis.

- *Scenario 1, no covariate effect:* we take C_i from $N(10,\ 2.5^2)$, with $\gamma = 0$.

- *Scenario 2, a common covariate effect:* we take C_i from $N(10,\ 2.5^2)$, with $\gamma = 1$.

- *Scenario 3, a covariate distribution that varies with treatment groups:* we take C_i from $N(13, 2.5^2)$ for $d = 1$, but from $N(10,\ 2.5^2)$ for $d = 0$, with $\gamma = 1$.

- *Scenario 4, a covariate distribution that has a scale change across treatment groups:* we take C_i from $N(10, 3^2)$ for $d = 1$, but from $N(10,\ 2.5^2)$ for $d = 0$, with $\gamma = 1$.

The choice of weights, $q_{d,i}$, assigns more weights to the overlap of the two groups of covariates. For each scenario, we partition the overlap, $I = [a_1, a_2]$, into three subintervals with equal length, I_1, I_2, and I_3. We assign weights 1 to the values of covariates outside the overlap, i.e., $(-\infty, a_1)$ and (a_2, ∞). Table 3.1 provides the weights for each scenario. The values of l and k are determined by (3.4).

Table 3.1: The choice of weights for all scenarios. The covariates from two groups overlap on $I = [a_1, a_2] = I_1 \cup I_2 \cup I_3$.

	values of covariates				
	$(-\infty, a_1)$	I_1	I_2	I_3	(a_2, ∞)
d=1	1	l^3	l^2	l	1
d=0	1	k	k^2	k^3	1

Type I Error

A total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 100$ are generated for Z under the null hypothesis (with $\eta = 0$). Table 3.2 shows that the Type I errors stay around the nominal level 5%, although q.COVES is a little liberal.

Table 3.2: Type I error at sample size $(n_1,\ n_0) = (100, 100)$ and nominal level 5% for i.i.d errors. The values in the table are in percentage.

Scenario	0.5th q.COVES test	0.75th q.COVES test	0.5th COVES test	0.75th COVES test
1	5.94	6.73	5.30	5.64
2	5.89	6.81	5.33	5.71
3	6.46	7.85	5.33	5.71
4	6.03	6.69	5.33	5.75
Scenario	0.5th rank test	0.75th rank test	t test	
1	4.96	5.05	5.19	
2	4.96	5.05	5.19	
3	4.76	5.14	5.37	
4	5.00	5.14	5.23	

Power Function of η

We compare power functions for these seven tests with η up to 1 (with grid 0.25) given sample size $n_1 = n_0 = 100$ and Type I error calibrated to the nominal level 5%. We use smoothing splines to obtain power curves. The results in Figure 3.1 clearly show the superiority of the COVES test.

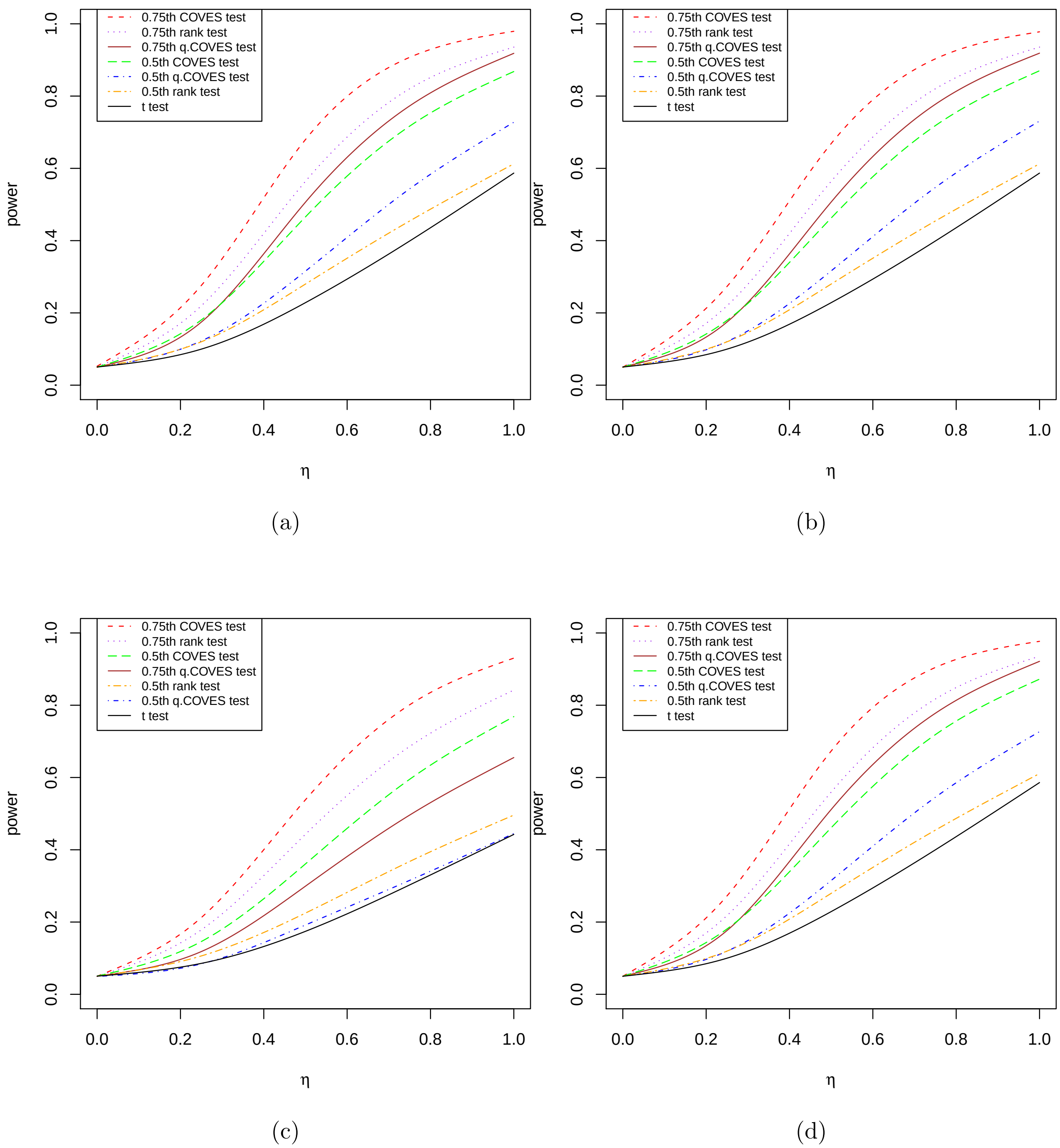


Figure 3.1: Statistical powers of seven tests as functions of η for i.i.d errors. (a) Scenario 1. (b) Scenario 2. (c) Scenario 3. (d) Scenario 4.

3.1.2 Simulation Studies for non-i.i.d Errors with Discrete Covariates

For simplicity, we assume discrete covariates first and then generalize to continuous covariates in the next section for non-i.i.d error models.

We consider data generated from (3.1) with four scenarios for the effects of the covariate in the analysis.

- *Scenario 1, no covariate effect:* we take C_i from $\{1, 2, 3\}$ with probability $\{0.2, 0.6, 0.2\}$ and $\gamma(u) = 0$.
- *Scenario 2, a common covariate effect:* we take C_i from $\{1, 2, 3\}$ with probability $\{0.2, 0.6, 0.2\}$ and $\gamma(u) = 1$.
- *Scenario 3, a covariate distribution that varies with treatment groups:* we take C_i from $\{2, 3, 10\}$ with probability $\{0.2, 0.5, 0.3\}$ for $d = 1$, but from $\{1, 2, 3\}$ with probability $\{0.2, 0.6, 0.2\}$ for $d = 0$. $\gamma(u) = 100u$.
- *Scenario 4, a covariate distribution that has a scale change across treatment groups:* we take C_i from $\{1, 2, 3\}$ with probability $\{0.2, 0.3, 0.5\}$ for $d = 1$, but from $\{2, 3, 4\}$ with probability $\{0.8, 0.1, 0.1\}$ for $d = 0$. $\gamma(u) = 100u$.

The choice of weights, $q_{d,i}$, assigns more weights to the common values of the two groups of covariates. For both groups, we assign weights 1 to the unique values in each of the two groups of covariates. Table 3.3 provides the weights for each scenario, where l and k are determined by (3.4). Note that, for Scenarios 1 and 2, we tried weights l^3, l^2, and l, and k, k^2, and k^3 for the values of covariates 1, 2, and 3 for $d = 1$ and $d = 0$, respectively, but no solutions to (3.4) were found.

Table 3.3: The choice of weights for each scenario.

	Scenario 1 and 2 values of covariates			Scenario 3 values of covariates				Scenario 4 values of covariates			
	1	2	3	1	2	3	10	1	2	3	4
d=1	l^4	l^3	l		l^2	l	1		l^2	l	1
d=0	k	k^3	k^4	1	k	k^2		1	k	k^2	

Type I Error

A total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 100$ are generated for Z under the null hypothesis (with $\delta(u) = 0$). From Table 3.4 and Figure 3.2, we find that the COVES test and the t test are not appropriate when the covariate between two groups differs in means. Heteroscedasticity is pronounced in scenario 3; therefore, the t test is not valid.

Table 3.4: Type I error at sample size $(n_1,\ n_0) = (100, 100)$ and nominal level 5% for non-i.i.d errors with discrete covariates. The values in the table are in percentage.

Scenario	0.5th q.COVES test	0.75th q.COVES test	0.5th COVES test	0.75th COVES test
1	5.88	6.35	5.12	5.38
2	5.99	6.41	5.32	5.52
3	5.14	5.49	**49.84**	**25.61**
4	5.31	5.47	4.55	5.77
Scenario	0.5th rank test	0.75th rank test	t test	
1	4.82	4.91	5.02	
2	4.82	4.91	5.02	
3	5.50	5.96	**0.43**	
4	5.41	5.30	5.48	

Power Function of a

We compare power functions for these seven tests with a up to 4 (with grid 0.5) for Scenario 1 and Scenario 2, and up to 200 (with grid 25) for Scenario 3 and Scenario 4, given the sample size $n_1 = n_0 = 100$, and Type I error calibrated to the nominal level 5%. We use smoothing splines to obtain power curves. The results in Figure 3.3 clearly show the superiority of the generalized COVES test in all scenarios.

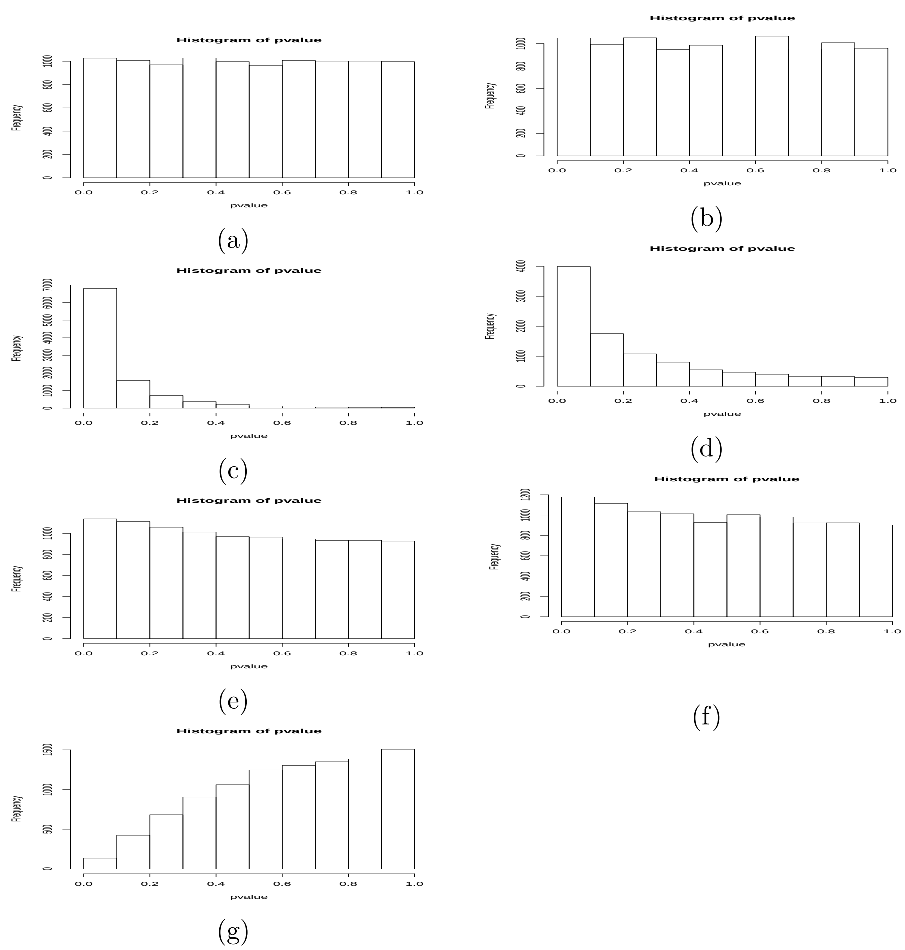
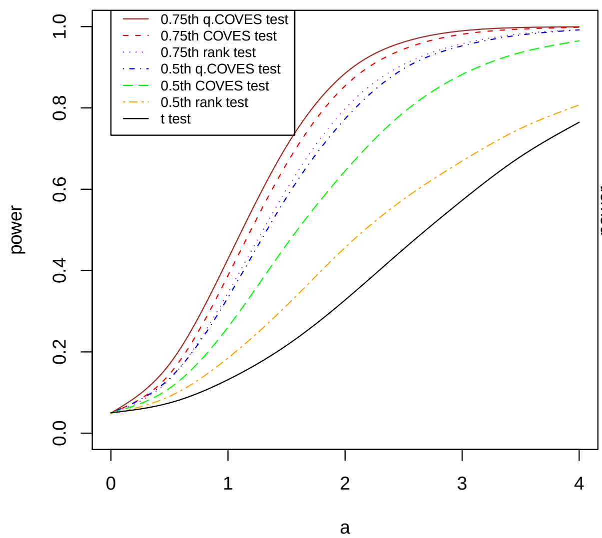
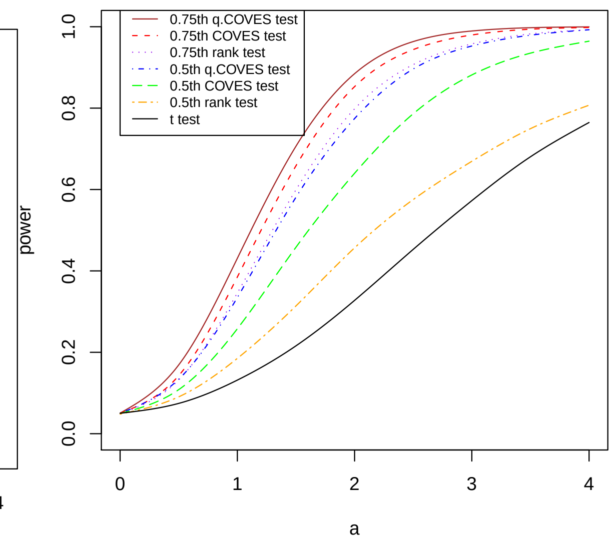


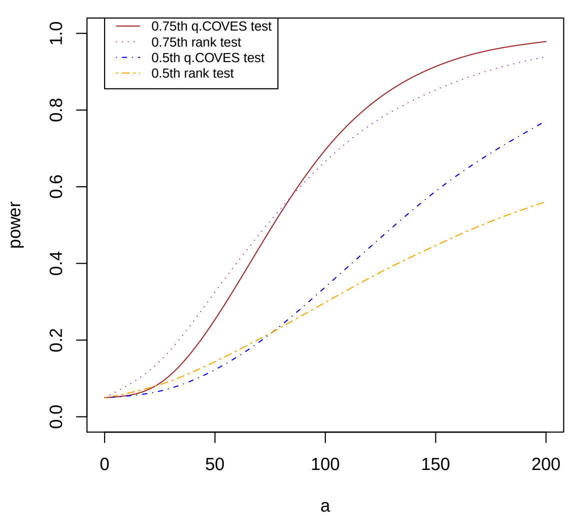
Figure 3.2: Histograms of p-values in Scenario 3 for non-i.i.d errors with discrete covariates. (a) 0.5th q.COVES test. (b) 0.75th q.COVES test. (c) 0.5th COVES test. (d) 0.75th COVES test. (e) 0.5th rank test. (f) 0.75th rank test. (g) t test.



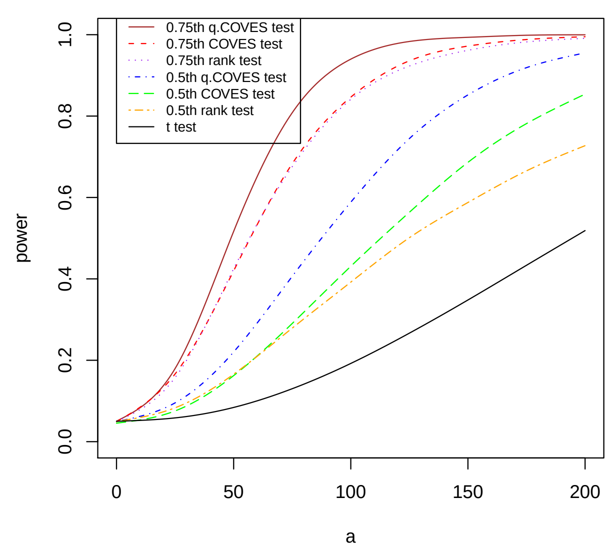
(a)



(b)



(c)



(d)

Figure 3.3: Statistical powers of seven tests as functions of a for non-i.i.d errors with discrete covariates. (a) Scenario 1. (b) Scenario 2. (c) Scenario 3. (d) Scenario 4.

3.1.3 Simulation Studies for non-i.i.d Errors with Normal Covariates

We consider data generated from (3.1) with four scenarios for the effects of the covariate in the analysis.

- *Scenario 1, no covariate effect:* we take C_i from $N(10,\ 2.5^2)$, with $\gamma(u) = 0$.
- *Scenario 2, a common covariate effect:* we take C_i from $N(10,\ 2.5^2)$, with $\gamma(u) = 1$.
- *Scenario 3, a covariate distribution that varies with treatment groups:* we take C_i from $N(13, 2.5^2)$ for $d = 1$, but from $N(10,\ 2.5^2)$ for $d = 0$, with $\gamma(u) = 100u$.
- *Scenario 4, a covariate distribution that has a scale change across treatment groups:* we take C_i from $N(10, 3^2)$ for $d = 1$, but from $N(10,\ 2.5^2)$ for $d = 0$, with $\gamma(u) = 100u$.

For each scenario, the choice of weights is the same as Table 3.1.

Type I Error

A total of $M = 10,000$ Monte Carlo samples of size $n_1 = n_0 = 100$ are generated for Z under the null hypothesis (with $\delta(u) = 0$). From Table 3.5 and Figure 3.4, we find that the COVES test is not appropriate when the covariate between two groups differs in means.

Table 3.5: Type I error at sample size $(n_1,\ n_0) = (100, 100)$ and nominal level 5% for non-i.i.d errors with normal covariates. The values in the table are in percentage.

Scenario	0.5th q.COVES test	0.75th q.COVES test	0.5th COVES test	0.75th COVES test
1	5.92	6.74	5.39	5.60
2	6.03	6.70	5.28	5.48
3	5.60	6.45	**17.90**	**7.14**
4	5.53	5.60	5.01	5.11

Scenario	0.5th rank test	0.75th rank test	t test
1	5.13	4.91	5.04
2	5.13	4.91	5.04
3	4.66	4.73	4.93
4	5.04	4.89	5.21

Power Function of a

We compare power functions for these seven tests with a up to 4 (with grid 0.5) for Scenario 1 and Scenario 2, and up to 800 (with grid 100) for Scenario 3 and Scenario 4, given the sample size $n_1 = n_0 = 100$, and Type I error calibrated to the nominal level 5%. We use smoothing splines to obtain power curves. The results in Figure 3.5 clearly show the superiority of the generalized COVES test in all scenarios.

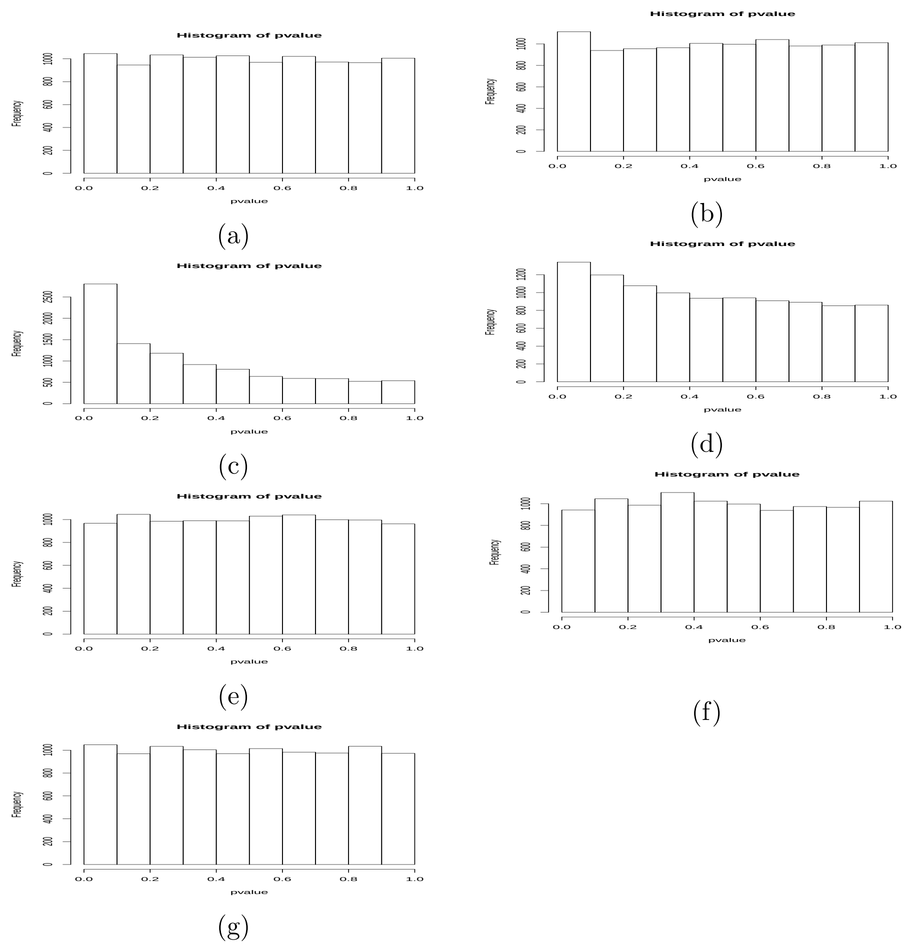
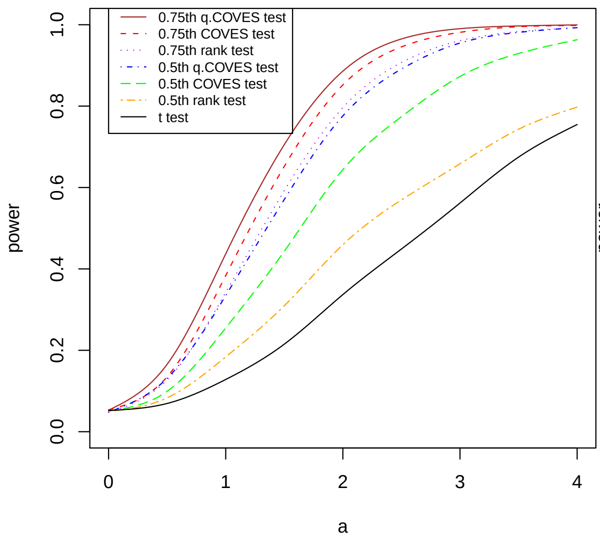


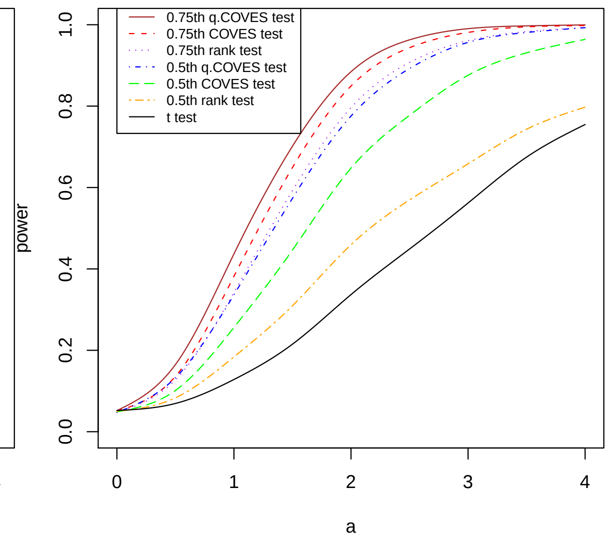
Figure 3.4: Histograms of p-values in Scenario 3 for non-i.i.d errors with normal covariates. (a) 0.5th q.COVES test. (b) 0.75th q.COVES test. (c) 0.5th COVES test. (d) 0.75th COVES test. (e) 0.5th rank test. (f) 0.75th rank test. (g) t test.

3.2 Future Work

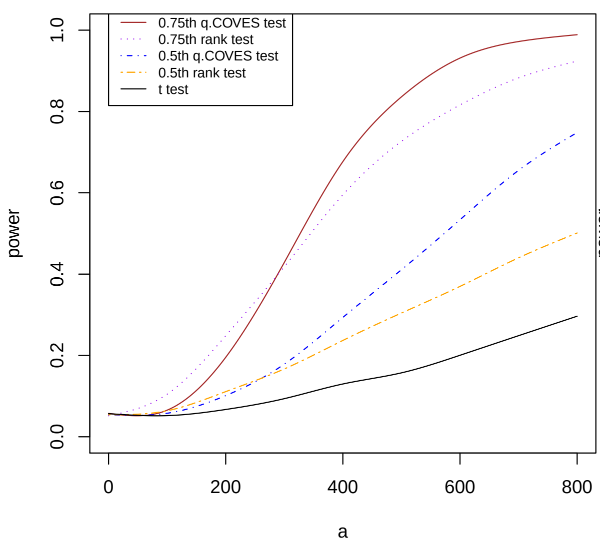
So far, we have provided a way to assign weights for the generalized covariate-adjusted expected shortfall test. As future work, we will investigate the optimal weights so that the q.COVES is as powerful as possible.



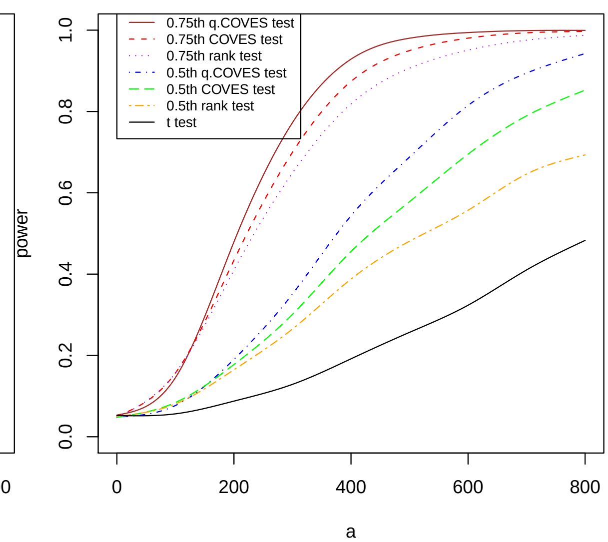
(a)



(b)



(c)



(d)

Figure 3.5: Statistical powers of seven tests as functions of a for non-i.i.d errors with noraml covariates. (a) Scenario 1. (b) Scenario 2. (c) Scenario 3. (d) Scenario 4.

Chapter 4

Bayesian Inference for Conditional Autoregressive Value at Risk

4.1 Introduction

The recent financial crisis has highlighted the importance of risk management. Value at Risk (VaR) is simply a quantile of the loss distribution of a portfolio within a given time period (k) and a confidence level ($1-\theta$). For example, at $k = 14$ days and the 0.99th quantile, if the 1% VaR is 5 million dollars, then there is 1% chance of a loss exceeding 5 million dollars over a period of two weeks. VaR becomes a popular measure of risk used by financial institutions because it is just a single number and is easy to understand. Many firms use overnight ($k = 1$ day) VaRs for internal management, and the two-week ($k = 14$ days) VaRs for disclosure to regulators. For example, J.P. Morgan discloses its daily 5% VaRs, and Bankers Trust discloses the daily 1% VaRs. Central bank regulators use VaR in determining the capital a bank is required to keep.

Accurate VaR estimation can help financial institutions maintain appropriate capital levels to cover the risk from the corresponding portfolio. A Variance-Covariance approach is often used, and is based on the normality assumption on the return distributions. Let $L_{t,k} = -y_{t,k}$ be the k-period loss, where $y_{t,k}$ is the k-period return defined as the sum of k period successive returns starting from time t, and the daily returns are computed as the difference of the log of successive prices. Assume that $y_{t,k}$ is normally distributed with mean $\mu_{t,k}$ and variance $\sigma^2_{t,k}$. According to the definition of VaR, $P(L_{t,k} > \mathrm{VaR}_{t,k,\theta}) = P(y_{t,k} < -\mathrm{VaR}_{t,k,\theta}) = \theta$, thus we have $\mathrm{VaR}_{t,k,\theta} = -(\mu_{t,k} + \Phi^{-1}(\theta)\sigma_{t,k})$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Assume that $\mu_{t,k} = 0$, and that the returns are uncorrelated with the same variance, then a k-period variance forecast $\hat{\sigma}^2_{t,k}$ is the sum of k period variances, i.e., $k\hat{\sigma}^2_{t+1}$. The Variance-Covariance approach is easy to implement because we just need to estimate the one-step-ahead variance forecast, $\hat{\sigma}^2_{t+1}$, by the Exponentially Weighted Moving Average (EWMA) model or the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model (Bollerslev, 1986).

However, Hull and White (1998) point out that the distributions of returns have positive excess kurtosis, i.e., the distributions have heavier tails than the normal distributions. Motivated by this observation, distribution-free approaches to estimate VaR seem to be more appropriate.

The quantile regression approach developed by Koenker and Bassett (1978) provides an estimate for a specific quantile on return. Suppose that we observe a vector $\{x_t,\ y_t\}_{t=1}^T$, where x_t is a p-dimensional explanatory variable, and y_t is the return at time t. Assume that the θ-th conditional quantiles of y_t given x_t is $x_t^T\beta(\theta)$, where $\theta \in (0,\ 1)$. Then, VaR at time t with a given confidence level $1-\theta$ and a holding period 1 day, $\mathrm{VaR}_{t,\theta}(\beta)$, equals $-x_t^T\beta(\theta)$. The θ-th regression quantile, $\hat{\beta}$, can be solved by a minimization problem

$$\min_{\beta} RQ(\beta, \theta),$$

where

$$RQ(\beta,\theta) = \sum_{t=1}^{T}\{\theta - I(y_t < -\mathrm{VaR}_{t,\theta}(\beta))\}\{y_t + \mathrm{VaR}_{t,\theta}(\beta)\} \tag{4.1}$$

is the quantile loss function, and I denotes the indicator function. Taylor (1999) presents a conditional quantile model to estimate the VaR over the holding period, k. Candidate explanatory variables are first chosen from any two products of $1, k, k^{1/2}, k^2, \hat{\sigma}_{t+1}$, and $\hat{\sigma}_{t+1}^2$, where $\hat{\sigma}_{t+1}$ is a GARCH(1, 1) (Bollerslev, 1986) one-step-ahead volatility forecast. He chooses explanatory variables with the largest Pseudo-R^2

$$\begin{aligned}\text{Pseudo-}R^2 &= 1 - \frac{\text{Sum of Weighted Deviations about Estimated Quantile}}{\text{Sum of Weighted Deviations about Raw Quantile}}\\ &= 1 - \frac{RQ(\hat{\beta},\theta)}{RQ(\theta)},\end{aligned} \tag{4.2}$$

where the raw quantile defined as $\hat{Q}(\theta)$ is the θ-th quantile of $\{y_t\}_{t=1}^T$, and $RQ(\theta) = \sum_{t=1}^{T}\{\theta - I(y_t < \hat{Q}(\theta))\}\{y_t - \hat{Q}(\theta)\}$.

Engle and Manganelli (2004) propose several conditional autoregressive value at risk (CAViaR) models. Consider the model for the return, y_t, at time period t, as

$$\begin{aligned} y_t &= Quant_\theta(y_{t-1}, x_{t-1}, \cdots, y_1, x_1; \beta) + \epsilon_{t,\theta}\\ &\equiv Quant_{t,\theta}(\beta) + \epsilon_{t,\theta}, \quad t = 1, \cdots, T,\end{aligned} \tag{4.3}$$

where x_t denotes a vector of explanatory variables, $Quant_{t,\theta}(\beta) \equiv Quant_\theta(y_{t-1}, x_{t-1}, \cdots, y_1, x_1; \beta)$

denotes the θ-th conditional quantile of return at time t, the θ-th conditional quantile of the error term given Ω_t is zero, i.e., $Quant_\theta(\epsilon_{t,\theta}|\Omega_t) = 0$, and $\Omega_t = (y_{t-1}, x_{t-1}, \cdots, y_1, x_1, Quant_{1,\theta}(\beta))$ is the information set available before time t.

Assume $\mathrm{VaR}_{t,\theta}(\beta)$ follows a generic conditional autoregressive value at risk (CAViaR) model

$$\mathrm{VaR}_{t,\theta}(\beta) = \beta_1 + \sum_{i=1}^{q} \beta_{i+1} \mathrm{VaR}_{t-i,\theta}(\beta) + \sum_{j=1}^{r} \beta_{q+j+1} l(x_{t-j}),$$

where $p = q + r + 1$ is the dimension of β, and l is a function of lagged explanatory variables. A natural choice of x_{t-j} is lagged returns.

Let $L_t = -y_t$ equal the loss at time t. Because $P(L_t > \mathrm{VaR}_{t,\theta}(\beta)) = P(y_t < -\mathrm{VaR}_{t,\theta}(\beta)) = \theta$, we have $\mathrm{VaR}_{t,\theta}(\beta) = -Quant_{t,\theta}(\beta)$ by (4.3).

Four specific CAViaR models used by Engle and Manganelli (2004) are:

Symmetric Absolute Value (SAV)

$$\mathrm{VaR}_{t,\theta}(\beta) = \beta_1 + \beta_2 \mathrm{VaR}_{t-1,\theta}(\beta) + \beta_3 |y_{t-1}|;$$

Asymmetric Slope (AS)

$$\mathrm{VaR}_{t,\theta}(\beta) = \beta_1 + \beta_2 \mathrm{VaR}_{t-1,\theta}(\beta) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^-,$$

where $(x)^+ = \max(x, 0)$, $(x)^- = -\min(x, 0)$;

Indirect GARCH(1, 1)

$$\mathrm{VaR}_{t,\theta}(\beta) = (\beta_1 + \beta_2 \mathrm{VaR}^2_{t-1,\theta}(\beta) + \beta_3 y^2_{t-1})^{1/2};$$

and Adaptive,

$$\mathrm{VaR}_{t,\theta}(\beta_1) = VaR_{t-1,\theta}(\beta_1) + \beta_1 [\{1 + \exp(G\{y_{t-1} + \mathrm{VaR}_{t-1,\theta}(\beta_1)\})\}^{-1} - \theta],$$

where G is some positive finite number. Generally, the parameter G could be estimated. For simplicity, Engle and Manganelli (2004) set $G = 10$.

All four models assume the dependence of VaR on the past return y_{t-1}. In the SAV and the indirect GARCH models, the past return is assumed to have a symmetric impact on the current VaR, i.e., past returns with the same magnitude but different sign will have the same impact on

VaR. On the other hand, the AS model allows an asymmetric impact.

All four models assume the dependence of VaR on the lagged VaR. In the previous three models, the corresponding coefficient of $\text{VaR}_{t-1,\theta}$ needs to be learned from the data, while the Adaptive model assumes a unit coefficient. The second term in the Adaptive model is a smoothed version of its own limit (when $G \to \infty$), a step function $\beta_1\{I(y_{t-1} \leq -\text{VaR}_{t-1,\theta}(\beta_1)) - \theta\}$. The intuition behind this model is that when the past loss is less than the past VaR, i.e., $y_{t-1} > -\text{VaR}_{t-1,\theta}(\beta_1)$, we should decrease the current VaR slightly by the amount of $\theta\beta_1$; when the past loss is exceeding or equal to the past VaR, i.e., $y_{t-1} \leq -\text{VaR}_{t-1,\theta}(\beta_1)$, we should increase the current VaR by the amount of $(1-\theta)\beta_1$.

Let Hits

$$\text{Hits}_t(\beta) \equiv I(y_t < -\text{VaR}_{t,\theta}(\beta)) - \theta. \tag{4.4}$$

Clearly, the expected value of $\text{Hits}_t(\beta)$ is 0. The θ-th regression quantile, $\hat{\beta}$, is solved by minimizing $RQ(\beta,\theta)$ in (4.1). Smaller values of $|\text{Hits}_t(\hat{\beta})|$ over a period of time are needed for $\hat{\beta}$ to be a good estimate. Engle and Manganelli (2004) estimate $\hat{\beta}$ by the Nelder-Mead simplex algorithm and the quasi-Newton method, and evaluate the performance of a CAViaR model by empirical Hits in-sample (Hits_{IS}) and empirical Hits out-of-sample (Hits_{OOS}).

This chapter contrasts the frequentist approach with a Bayesian approach on the CAViaR models for providing the entire posterior distributions of interests in parameter, VaR, Hits in-sample, and Hits out-of-sample. Engle and Manganelli's method of estimation may not find the optimal quantile estimates that minimize the quantile loss function (4.1) because smaller values were attained by the Bayesian approach. Importantly, the performance of CAViaR models is better using the Bayesian approach than Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria. The remainder of this chapter is organized as follows. We provide the working likelihood for the error term, the prior, and the proof of the property of the posterior. Next, we propose an MCMC strategy along with a block algorithm for the Bayesian inference. In the end, we present empirical investigations in both a stable period and volatile period to show that the Bayesian approach adds value to the original estimation method of Engle and Manganelli (2004) in terms of both parameter estimation and model fitting.

4.2 Bayesian Modeling for VaR

In this Section, we provide the Bayesian settings on the CAViaR models.

4.2.1 Working Likelihood

Consider an asymmetric Laplace density as a working conditional density function for the error term

$$f(\epsilon_{t,\theta}|\Omega_t) = \frac{\theta(1-\theta)}{\sigma} \exp\{-\frac{1}{\sigma}\rho_\theta(\epsilon_{t,\theta})\},$$

where $\rho_\theta(\epsilon_{t,\theta}) = \epsilon_{t,\theta}\{\theta - I(\epsilon_{t,\theta} < 0)\}$ is a loss function, and σ is a scale parameter.

Then the working likelihood can be expressed as

$$f(\text{data}|\beta,\sigma) = \frac{\theta^T(1-\theta)^T}{\sigma^T} \exp(-\frac{1}{\sigma}RQ(\beta,\theta)), \tag{4.5}$$

where data denotes $(y_1, \ldots, y_T)$, and $RQ(\beta,\theta)$ is defined in (4.1).

The appealing property of this working likelihood is that the corresponding maximum likelihood estimator for β is equivalent to the quantile regression estimator, i.e., the one minimizing $RQ(\beta,\theta)$ as provided in Engle and Manganelli (2004). An alternative to this parametric working likelihood is to model the unknown error distribution nonparametrically, such as the Dirichlet process mixture models developed by Kottas and Krnjajić (2009).

4.2.2 Prior and Posterior Specification

We choose a flat prior for each coefficient, $\beta_i \in B_i$, for some interval B_i with $i = 1, \cdots, p$, and the Inverse Gamma distribution IG(α_0, s_0) for σ. We assume the impact of the lagged VaR is bounded, and specify Uniform$[-L, L]$ as the prior distribution for β_2, the coefficient associated with VaR$_{t-1,\theta}(\beta)$, where L is a positive number. Assume independent components, the joint prior for all parameters is given by

$$\pi(\beta,\sigma) \propto \frac{1}{\sigma^{\alpha_0+1}} \exp(-\frac{s_0}{\sigma}) \text{ for } \beta_i \in B_i,\ \sigma > 0. \tag{4.6}$$

The posterior for (β, σ) can be derived as

$$\begin{aligned} f(\beta, \sigma|\text{data}) &\propto f(\text{data}|\beta, \sigma)\pi(\beta, \sigma) \\ &= \frac{\theta^T(1-\theta)^T}{\sigma^{T+\alpha_0+1}} \exp(-\frac{1}{\sigma} RQ(\beta, \theta) - \frac{s_0}{\sigma}) \text{ for } \beta_i \in B_i,\ \sigma > 0. \end{aligned}$$

Yu and Moyeed (2001) proved that the posterior is proper when VaR is a linear function of some explanatory variables x_t, i.e., $\text{VaR}_{t,\theta}(\beta) = -x_t^T \beta(\theta)$. We will prove that the posterior is proper for CAViaR models, and all posterior moments for β exist under some mild conditions.

Lemma 1 of Yu and Stander (2007)

Let the function $g(t) = \exp(-|t|)$, and $f(t) = \theta(1-\theta)\exp(-t[\theta - I(t<0)])$, then $f(t)$ has upper bound $g(h_1(\theta))$ and lower bound $g(h_2(\theta)t)$.

Theorem 4.2.1. *If the likelihood is given by (4.5) and the prior distribution is given by (4.6), then the corresponding posterior distribution is proper when $T \geq m$, where m is the dimension of improper prior for β. That is,*

$$\int_0^\infty \int_{\boldsymbol{B}} f(\beta, \sigma|data) d\beta d\sigma < \infty, \tag{4.7}$$

where $\boldsymbol{B} = B_1 \times \cdots B_p$, p is the dimension of β, $B_2 = (-L, L)$ for some large number L, $B_i = (-\infty, \infty)$ for $i = 1, 3, \cdots, p$, $p = 3$ and $m = 2$ for the SAV and the indirect GARCH models, $p = 4$ and $m = 3$ for the AS model, and $p = m = 1$ for the Adaptive model. Further, the following posterior moments of β,

$$\int_0^\infty \int_{\boldsymbol{B}} |\beta_1|^{r_1} \cdots |\beta_p|^{r_p} f(\beta, \sigma|data) d\beta d\sigma, \tag{4.8}$$

exist if the sample size $T > m + r_1 + r_3 + \cdots + r_p - \alpha_0$, where $(r_1, \cdots, r_p)$ denotes the order of the moments for each element of β.

Proof of Theorem 4.2.1 We consider the SAV model first. For simplicity, let $\text{VaR}_{t,\theta}(\beta) = V_t$. Expand the SAV model, we have

$$\begin{aligned} V_t &= \beta_1(1 + \beta_2 + \cdots + \beta_2^{t-1}) + \beta_2^t V_0 + \beta_3(|y_{t-1}| + \beta_2|y_{t-2}| + \cdots + \beta_2^{t-1}|y_0|) \\ &= \beta_1(1 + \beta_2 + \cdots + \beta_2^{t-1}) + \beta_3(|y_{t-1}| + \beta_2|y_{t-2}| + \cdots + \beta_2^{t-2}|y_1|), \end{aligned}$$

where $y_0 = V_0 = 0$ WLOG. Note that, the SAV model does not explode if $|\beta_2| < L$, for some large L. Therefore, we have $B_1 = B_3 = (-\infty, \infty)$, and $B_2 = (-L, L)$. According to Lemma 1 of Yu and

Stander (2007), we have

$$\begin{aligned} f(\frac{y_t+V_t}{\sigma}) &= \theta(1-\theta)\exp[-\frac{1}{\sigma}\{\theta - I(y_t+V_t<0)\}(y_t+V_t)] \\ &= g(h(\theta)\frac{y_t+V_t}{\sigma}) \\ &= \exp(-\frac{|h(\theta)|}{\sigma}|y_t+V_t|). \end{aligned}$$

Therefore, the posterior distribution

$$f(\beta,\sigma|\text{data}) = \frac{1}{\sigma^{T+\alpha_0+1}}\exp(-\frac{s_0}{\sigma} - \frac{|h(\theta)|}{\sigma}\sum_{t=1}^{T}|y_t+V_t|).$$

It suffices to show that for any $r_1,\ r_2\ r_3 \geq 0$, the posterior moment for β

$$\begin{aligned} &\int_0^\infty \int_{\boldsymbol{B}} \prod_{i=1}^{p} |\beta_i|^{r_i} f(\beta,\sigma|\text{data}) d\beta d\sigma \\ \leq\ &\int_0^\infty \frac{e^{-s_0/\sigma}}{\sigma^{T+\alpha_0+1}} \int_{-L}^{L} |\beta_2|^{r_2} \int_{\mathbb{R}^2} |\beta_1|^{r_1}|\beta_3|^{r_3} \exp(-\frac{|h(\theta)|}{\sigma}\sum_{t=1}^{2}|y_t+V_t|) d\beta_1 d\beta_3 d\beta_2 d\sigma \end{aligned} \tag{4.9}$$

is finite. Change variables (β_1,β_3) to (u,v) with

$$\begin{aligned} u &= |h(\theta)|(y_1+V_1)/\sigma = |h(\theta)|(\beta_1+y_1)/\sigma \\ v &= |h(\theta)|(V_2+y_2)/\sigma = |h(\theta)|(\beta_1+\beta_1\beta_2+\beta_3|y_1|+y_2)/\sigma, \end{aligned}$$

we have

$$\begin{aligned} \beta_1 &= \frac{\sigma}{|h(\theta)|}u - y_1, \\ \beta_3 &= \frac{-(1+\beta_2)\sigma}{|h(\theta)y_1|}u + \frac{\sigma}{|h(\theta)y_1|}v + \frac{(1+\beta_2)y_1-y_2}{|y_1|}, \\ J &= \begin{vmatrix} \frac{\sigma}{|h(\theta)|} & 0 \\ \frac{-(1+\beta_2)\sigma}{|h(\theta)y_1|} & \frac{\sigma}{|h(\theta)y_1|} \end{vmatrix} = \frac{\sigma^2}{h(\theta)^2|y_1|}. \end{aligned}$$

Then the integral (4.9) can be re-written as

$$\begin{aligned} &\frac{1}{|h(\theta)|^{T+\alpha_0}|y_1|^{r_3+1}} \int_0^\infty a^{T+\alpha_0-r_1-r_3-3}\exp(-\frac{s_0}{|h(\theta)|}a) \int_{-L}^{L}|\beta_2|^{r_2}\int_{\mathbb{R}^2}|u-ay_1|^{r_1} \\ &|-(1+\beta_2)u+v+a\{(1+\beta_2)y_1-y_2\}|^{r_3}\exp(-|u|-|v|)dudvd\beta_2 da, \end{aligned} \tag{4.10}$$

where we change variable σ to $a = |h(\theta)|/\sigma$. Let I be the integration over (u, v), $u = r\cos\theta$, $v = r\sin\theta$, we have

$$
\begin{aligned}
&|u - ay_1| = |r\cos\theta - ay_1| \leq r|\cos\theta| + a|y_1| \leq r + a|y_1|, \\
&|-(1+\beta_2)u + v + a\{(1+\beta_2)y_1 - y_2\}|^2 \leq 2(|-(1+\beta_2)u + v|^2 + |a\{(1+\beta_2)y_1 - y_2\}|^2) \\
&\leq 2(((1+\beta_2)^2 + 1)(u^2 + v^2) + a\{(1+\beta_2)y_1 - y_2\}^2) \text{ (Cauchy-Schwartz Inequality)} \\
&= 2((-(1+\beta_2)^2 + 1)r^2 + a\{(1+\beta_2)y_1 - y_2\}^2), \\
&-|u| - |v| = -r(|\cos\theta| + |\sin\theta|) \leq -r.
\end{aligned}
$$

Therefore, we have

$$
I \quad \leq \quad 2\pi \int_0^\infty (r + a|y_1|)^{r_1} [2\{(-(1+\beta_2) + 1)r^2 + a\{(1+\beta_2)y_1 - y_2\}^2\}]^{\frac{r_3}{2}} r \exp(-r) dr,
$$

which returns polynomial terms of a and β_2. Due to the boundedness of β_2, integration (4.10) is finite as long as $T > 2 + r_1 + r_3 - \alpha_0$. Specially T needs to at least $m = 2$ to have a proper posterior distribution. $\diamond$

4.3 Posterior Inference

In this section, we propose a fully Bayesian approach for the estimation and inference on the parameter β of the CAViaR models.

4.3.1 Element-wise Algorithm

We employ the Gibbs sampler for posterior sampling for the $p+1$ dimensional parameters (β, σ). To implement the Gibbs sampler, we need to derive the conditional distributions for each parameter given all the other parameters and the data:

$$
\begin{aligned}
&f(\beta_i|\beta_{[-i]}, \sigma, \text{data}) \propto f(\text{data}|\beta, \sigma)\pi(\beta_i), \; i = 1, \cdots, p, \text{ and} \\
&\sigma|\beta, \text{data} \sim IG(T + \alpha_0, RQ(\beta, \theta) + s_0),
\end{aligned}
$$

where $\beta_{[-i]} = (\beta_1, \cdots, \beta_{i-1}, \beta_{i+1}, \cdots, \beta_p)$, and IG stands for the Inverse Gamma distribution with shape parameter $T + \alpha_0$ and scale parameter $RQ(\beta, \theta) + s_0$.

The posterior distribution for each β_i cannot be reduced analytically to a well known distribution, so we used a Metropolis within Gibbs sampling to obtain samples of the model parameters from the posterior distribution.

Details of the element-wise algorithm are as follows.

Initialization: Choose starting values (β^0, σ^0) as $\beta_i^0 \sim U(0,1),\ i = 1, \cdots, p$ and $\sigma^0 \sim U(0,1)$.

Repeat the following sampling steps M times to obtain a chain for the parameters (β, σ):

Step i $(i = 1, \cdots, p)$: Sample $\tilde{\beta}_i^t$ by the Random-Walk Metropolis algorithm (Metropolis et al., 1953) using the proposal distribution

$$p(\beta_i^t | \beta_1^t, \cdots, \beta_{i-1}^t, \beta_{i+1}^{t-1}, \cdots, \beta_p^{t-1}, \sigma^{t-1}, \text{data}) = N(\beta_i^{t-1}, \nu^2),\ i = 1, \cdots, p,$$

where ν is a tuning parameter, which will be discussed later. The corresponding Metropolis ratio is given by

$$\begin{aligned} r_i &= \frac{f(\tilde{\beta}_i^t | \beta_1^t, \cdots, \beta_{i-1}^t, \beta_{i+1}^{t-1}, \cdots, \beta_p^{t-1}, \sigma^{t-1}, \text{data})}{f(\beta_i^{t-1} | \beta_1^t, \cdots, \beta_{i-1}^t, \beta_{i+1}^{t-1}, \cdots, \beta_p^{t-1}, \sigma^{t-1}, \text{data})} \\ &= \exp\{-\frac{1}{\sigma^{t-1}}(RQ(\tilde{\beta}_i^t, \theta) - RQ(\beta_i^{t-1}, \theta))\}, \end{aligned}$$

where $\tilde{\beta}_i^t = (\beta_1^t, \cdots, \beta_{i-1}^t, \tilde{\beta}_i^t, \beta_{i+1}^{t-1}, \cdots, \beta_p^{t-1})$, and $\beta_i^{t-1} = (\beta_1^t, \cdots, \beta_{i-1}^t, \beta_i^{t-1}, \beta_{i+1}^{t-1}, \cdots, \beta_p^{t-1})$. Draw $u_i \sim U(0,1)$, and set β_i^t to be

$$\beta_i^t = \begin{cases} \tilde{\beta}_i^t, \text{ if } u_i < min\{r_i, 1\}, \\ \beta_i^{t-1}, \text{ otherwise.} \end{cases}$$

Step $(p+1)$: Sample σ from $IG(T + \alpha_0, RQ(\beta, \theta) + s_0)$.

4.3.2 Case Study

In the element-wise MCMC implementation, we assume independent flat priors for all the coefficients on the CAViaR models and update one parameter at a time. To demonstrate a problem with the MCMC chain, we chose a sample of 3,392 daily prices from April 7, 1986 to April 7, 1999 for General Motors (GM) as Engle and Manganelli (2004) did. The first 2,892 (T) daily returns are treated as the training data or in-sample observations to estimate the model. We chose the starting values (β^0, σ^0) as $\beta_i^0 \sim U(0,1),\ i = 1, \cdots, p$ and $\sigma^0 \sim U(0,1)$, the tuning parameter, ν, to be 0.05, and small α_0 and s_0. It took about 11.2 minutes to obtain a chain of length 60,000 for the four parameters (three coefficients and one scale parameter) in the 1% SAV model (2.2 GHZ and 2 GB of

RAM). Summary statistics (mean, median, and standard deviation) are calculated to determine the burn-in period. According to Figure 4.1, all summary statistics are stable after 50,000. Therefore, we choose the burn-in period to be 50,000.

We use acceptance rate (AR) to check efficiency of the element-wise algorithm. Gelman et al. (1996) suggest the cutoff acceptance rate $0.3/d$, where d is the dimension of parameters. In the simulation, AR are 0.1995, 0.0518, and 0.1642 for β_1, β_2, and β_3, respectively for a chain with length 60,000. Because AR for β_2 is too low (less than $0.3/3 = 0.1$), the chain is not efficient. We will use a block algorithm to improve the efficiency in the next section.

4.4 Block Algorithm

It is more efficient to use a block algorithm to sample some coefficients together. This will be referred to as a preliminary run to sample intercept, β_1, and the remaining coefficients, $\beta_{[-1]} = (\beta_2, \cdots, \beta_p)$, respectively.

4.4.1 Sampling Schemes for Preliminary Run

Details of the preliminary run are as follows.

Initialization: Choose starting values (β^0, σ^0) as $\beta_i^0 \sim U(0,1),\ i = 1, \cdots, p$ and $\sigma^0 \sim U(0,1)$.

Repeat the following sampling steps M times to obtain a chain for the parameters (β, σ):

Step 1: Sample $\tilde{\beta}_1^t$ by the Random-Walk Metropolis algorithm with the proposal distribution

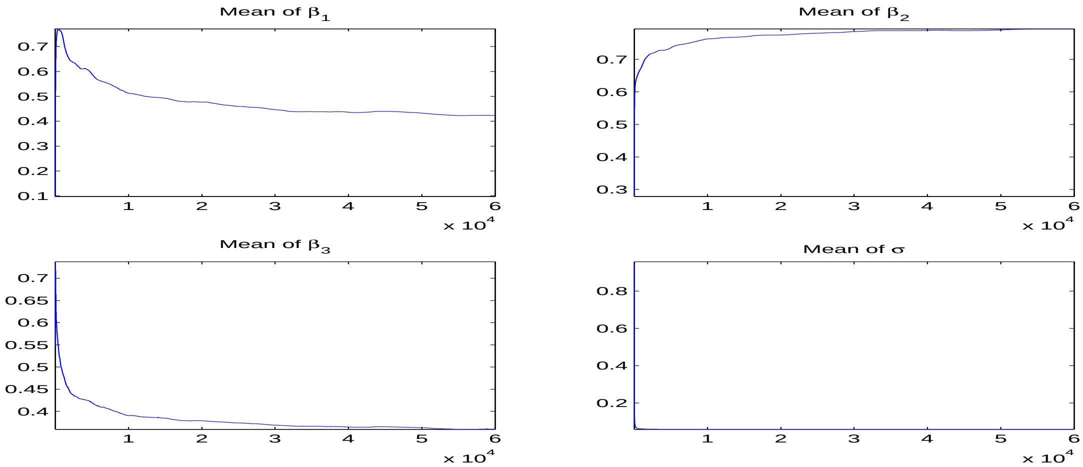
$$p(\beta_1^t|\beta_{[-1]}^{t-1}, \sigma^{t-1}, \text{data}) = N(\beta_1^{t-1}, \nu_1^2),$$

where ν_1 is a tuning parameter, and the target distribution is

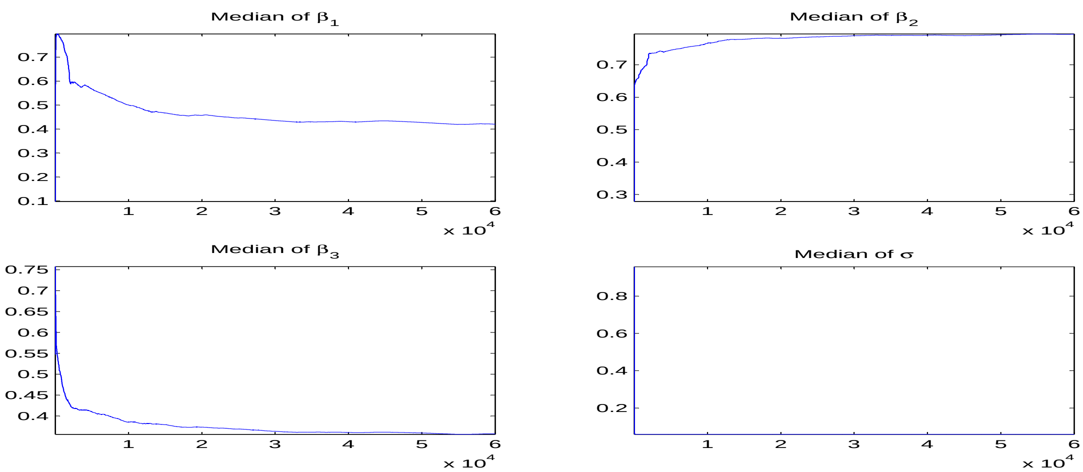
$$f(\beta_1^t|\beta_{[-1]}^{t-1}, \sigma^{t-1}, \text{data}) \propto \frac{1}{(\sigma^{t-1})^T} \exp(-\frac{1}{\sigma^{t-1}} RQ(\beta, \theta)).$$

The corresponding Metropolis ratio is given by

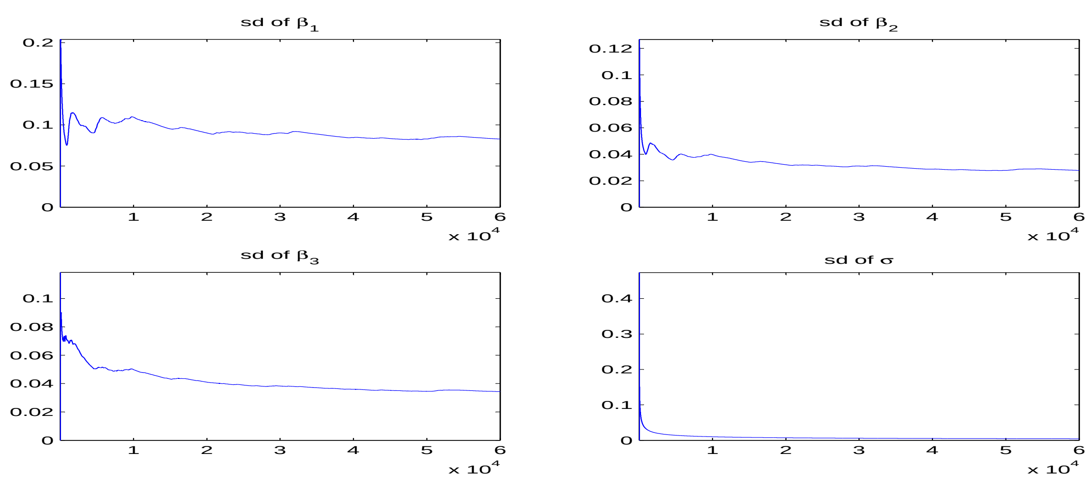
$$\begin{aligned} r_1 &= \frac{f(\tilde{\beta}_1^t|\beta_{[-1]}^{t-1}, \sigma^{t-1}, \text{data})}{f(\beta_1^{t-1}|\beta_{[-1]}^{t-1}, \sigma^{t-1}, \text{data})} \\ &= \exp\{-\frac{1}{\sigma^{t-1}}(RQ(\tilde{\beta}^t, \theta) - RQ(\beta^{t-1}, \theta))\}, \end{aligned}$$



(a)



(b)



(c)

Figure 4.1: Summary statistics for each parameter in the element-wise MCMC implementation: (a) mean . (b) median. (c) standard deviation.

where $\tilde{\beta}^t = (\tilde{\beta}_1^t, \beta_{[-1]}^{t-1})$, and $\beta^{t-1} = (\beta_1^{t-1}, \beta_{[-1]}^{t-1})$. Draw $u_1 \sim U(0,1)$, and set β_1^t to be

$$\beta_1^t = \begin{cases} \tilde{\beta}_1^t, \text{ if } u_1 < min\{r_1, 1\}, \\ \beta_1^{t-1}, \text{ otherwise.} \end{cases}$$

Step 2: Sample $\tilde{\beta}_{[-1]}^t$ by the Multivariate Random-Walk Metropolis algorithm with the proposal distribution

$$p(\beta_{[-1]}^t | \beta_1^t, \sigma^{t-1}, \text{data}) = N(\beta_{[-1]}^{t-1}, \Sigma_0),$$

where $\Sigma_0 = \nu_2^2 R$ is a variance-covariance matrix, ν_2 is a tuning parameter, and R is a correlation coefficient matrix with diagonal 1. We do not need to implement Step 2 for the Adaptive model because only β_1 exists in that model. The target distribution is

$$f(\beta_{[-1]}^t | \beta_1^t, \sigma^{t-1}, \text{data}) \propto \frac{1}{(\sigma^{t-1})^T} \exp(-\frac{1}{\sigma^{t-1}} RQ(\beta, \theta)).$$

The corresponding Metropolis ratio is given by

$$\begin{aligned} r_2 &= \frac{f(\tilde{\beta}_{[-1]}^t | \beta_1^t, \sigma^{t-1}, \text{data})}{f(\beta_{[-1]}^{t-1} | \beta_1^{t-1}, \sigma^{t-1}, \text{data})} \\ &= \exp\{-\frac{1}{\sigma^{t-1}} (RQ(\tilde{\beta}^t, \theta) - RQ(\beta^{t-1}, \theta))\}, \end{aligned}$$

where $\tilde{\beta}^t = (\beta_1^t, \tilde{\beta}_{[-1]}^t)$, $\beta^{t-1} = (\beta_1^t, \beta_{[-1]}^{t-1})$. Draw $u_2 \sim U(0,1)$, and set $\beta_{[-1]}^t$ to be

$$\beta_{[-1]}^t = \begin{cases} \tilde{\beta}_{[-1]}^t, \text{ if } u_2 < min\{r_2, 1\}, \\ \beta_{[-1]}^{t-1}, \text{ otherwise.} \end{cases}$$

Step 3: Sample σ from $IG(T + \alpha_0, RQ(\beta, \theta) + s_0)$.

In Step 1, we sample $\tilde{\beta}_1^t$ by the Random-Walk Metropolis algorithm. On the other hand, we can draw $\tilde{\beta}_1^t$ from the neighborhood of the mode of the target distribution, i.e., from the proposal distribution

$$p(\beta_1^t | \beta_{[-1]}^{t-1}, \sigma^{t-1}, \text{data}) = N(\text{mode}, \nu_1^2),$$

where the mode can be solved by a maximization problem

$$\max_{\beta_1^t} f(\beta_1^t|\beta_{[-1]}^{t-1}, \sigma^{t-1}, \text{data}).$$

This is equivalent to solve a minimization problem

$$\min_{\beta_1^t} RQ((\beta_1^t, \beta_{[-1]}^{t-1}), \theta).$$

For the SAV model

$$RQ((\beta_1^t, \beta_{[-1]}^{t-1}), \theta) = \sum_{t=1}^{T}\{(z_t + \beta_1^t)(\theta - I(z_t + \beta_1^t < 0))\},$$

where $z_t = y_t + \beta_2^{t-1}\text{VaR}_{t-1,\theta}((\beta_1^{t-1}, \beta_{[-1]}^{t-1})) + \beta_3^{t-1}|y_{t-1}|$, the mode is equal to (-1) times the θth quantile of $\{z_t\}_{t=1}^T$.

For the AS model

$$RQ((\beta_1^t, \beta_{[-1]}^{t-1}), \theta) = \sum_{t=1}^{T}\{(z_t + \beta_1^t)(\theta - I(z_t + \beta_1^t < 0))\},$$

where $z_t = y_t + \beta_2^{t-1}\text{VaR}_{t-1,\theta}((\beta_1^{t-1}, \beta_{[-1]}^{t-1})) + \beta_3^{t-1}(y_{t-1})^+\beta_4^{t-1}(y_{t-1})^-$, the mode is equal to (-1) times the θth quantile of $\{z_t\}_{t=1}^T$.

For the indirect GARCH model, because $RQ((\beta_1^t, \beta_{[-1]}^{t-1}), \theta)$ is a nonlinear function of β_1^t, we will stick with the Random-Walk Metropolis algorithm.

For the Adaptive model,

$$RQ(\beta_1^t, \theta) = \sum_{t=1}^{T}\{(z_t + w_t\beta_1^t)(\theta - I(z_t + \beta_1^t w_t < 0))\},$$

where $z_t = y_t + \text{VaR}_{t-1,\theta}(\beta_1^{t-1})$, and $w_t = \{1 + \exp(G\{y_{t-1} + \text{VaR}_{t-1,\theta}(\beta_1^{t-1})\})\}^{-1} - \theta$. Therefore, the mode, $\hat{\beta}_1(\theta)$, is the θth regression quantile of the above RQ function.

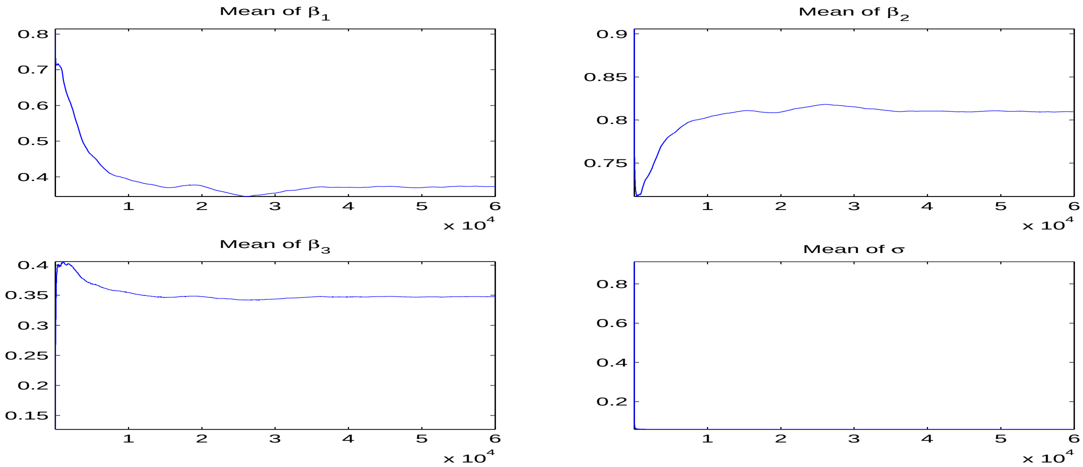
4.4.2 Case Study (Continued)

Using the same data analyzed in Subsection 4.3.2, we chose the starting values (β^0, σ^0) as $\beta_i^0 \sim U(0,1),\ i = 1, \cdots, p$ and $\sigma^0 \sim U(0,1)$, the tuning parameter, ν_1, and ν_2 to be 0.05, and small α_0 and s_0. Define R_{ij} =cor$(\beta_{i+1},\ \beta_{j+1})$ under the posterior. We set R_{12} to be negative in the SAV and indirect GARCH models. Besides, we set R_{12}, R_{13}, and R_{23} to be negative, negative, and positive, respectively, in the AS model. These settings are consistent to the estimated correlation matrix calculated from the independent posterior sequence generated from the element-wise MCMC implementation (burn-in=50,000 and lag=1,000) on each CAViaR model. In the SAV and indirect GARCH models, $R_{12} = -0.8037$ and -0.8592, respectively. Besides, in the AS model, $R_{12} = -0.8978$, $R_{13} = -0.8015$, and $R_{23} = 0.8020$. In addition, we set the off-diagonal, R_{ij}, to be 0.5 as the initial guess. The choice of those specific values may not be efficient; therefore, we will update the variance-covariance structure, Σ_0, used in the proposal distribution based on the data information in the second run.

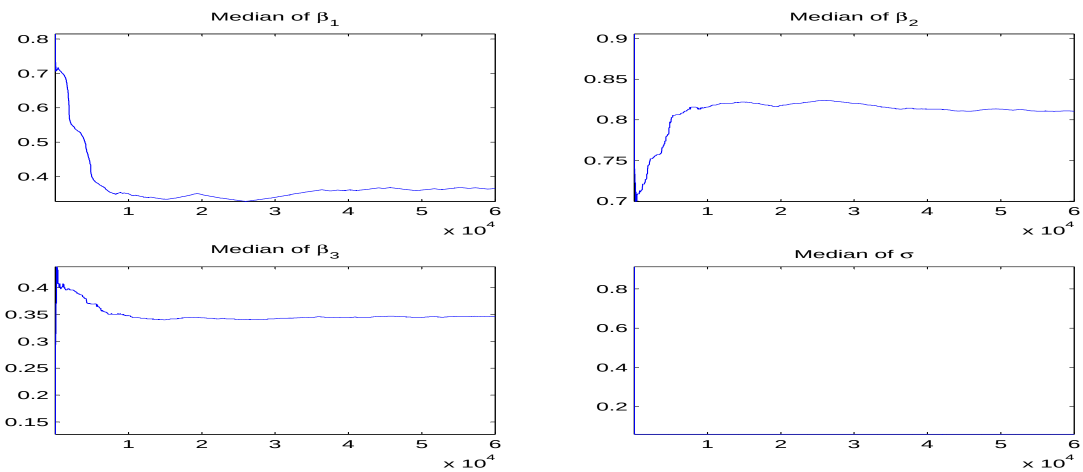
For the preliminary run, it took about 8.6 minutes to get a chain of length 60,000 for the four parameters (three coefficients and one scale parameter) in the 1% SAV model (2.2 GHZ and 2 GB of RAM). The AR values are 0.1885 and 0.0322 for β_1 and $\beta_{[-1]}$, respectively. The AR values decrease a little bit from 0.0518 and 0.1642 for β_2 and β_3 in the element-wise MCMC implementation to 0.0322 for $\beta_{[-1]}$ in the preliminary run using block algorithm. The main reason is the choice of the variance-covariance matrix, Σ_0. Summary statistics (mean, median, and standard deviation) are calculated to determine the burn-in period. According to Figure 4.2, all summary statistics are stable after 50,000. Therefore, we choose the burn-in period to be 50,000.

We take lag=1,000 to get nearly independent sequences of $\beta_{[-1]}$ to update the variance-covariance matrix used in the proposal distribution for $\beta_{[-1]}$. Besides, using the posterior sequences in the preliminary run, we found that the correlation of the intercept and each slope coefficient is high (-0.9 for correlation between β_1 and β_2 and 0.6 for correlation between β_1 and β_3). This indicates that the simulation is not efficient because posterior sequences do not move in the high density area. We suggest to perform reparameterization on the coefficients, i.e., centering each explanatory variable, to increase the efficiency, and then perform block algorithm on this transformed model. This will be called a second run. In other words, for the SAV, AS, and indirect GARCH models, we need to perform the second run on the transformed model using the updated variance-covariance matrix in the proposal distribution of $\beta_{[-1]}$ to increase the efficiency of the chain.

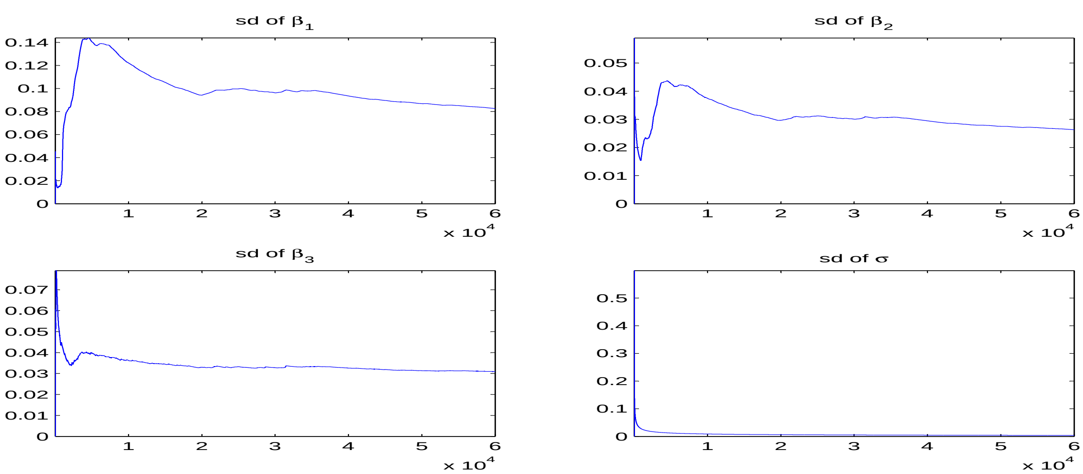
In addition, we perform the same sampling schemes except for Step 1, with $\tilde{\beta}_1^t$ sampled from the neighborhood of the mode of the target distribution. It took about 10.1 minutes to get a chain of length 60,000 for the four parameters (three coefficients and one scale parameter) in the 1% SAV model. The AR values are 0.1608 and 0.0399 for β_1 and $\beta_{[-1]}$, respectively. Results show that AR does not increase much over the Random-Walk Metropolis method.



(a)



(b)



(c)

Figure 4.2: Summary statistics for each parameter in the preliminary run: (a) mean . (b) median. (c) standard deviation.

4.4.3 Sampling Schemes for Second Run

Details of the second run are as follows.

Initialization: Do reparameterization on the CAViaR models by centering each explanatory variable, we get the new intercept β_1^*.

Symmetric Absolute Value (SAV)

$$\beta_1^* = \beta_1 + \beta_2 \overline{\mathrm{VaR}_{t-1,\ \theta}(\beta)} + \beta_3 \overline{|y_{t-1}|}, \tag{4.11}$$

Asymmetric Slope (AS)

$$\beta_1^* = \beta_1 + \beta_2 \overline{\mathrm{VaR}_{t-1,\ \theta}(\beta)} + \beta_3 \overline{(y_{t-1})^+} + \beta_4 \overline{(y_{t-1})^-}, \tag{4.12}$$

and Indirect GRACH(1, 1)

$$\beta_1^* = \beta_1 + \beta_2 \overline{\mathrm{VaR}_{t-1,\ \theta}(\beta)^2} + \beta_3 \overline{y_{t-1}^2}, \tag{4.13}$$

where overline denotes the average over $t = 2, \cdots, T$, and $\overline{\mathrm{VaR}_{t-1,\ \theta}(\beta)}$ and $\overline{\mathrm{VaR}_{t-1,\ \theta}(\beta)^2}$ are calculated using the posterior mean of β on the CAViaR model in the preliminary run.

Repeat the following sampling steps M times to obtain a chain for the parameters (β, σ):

Step 1: Choose starting values (β^{*0}, σ^0) as the posterior mean in the preliminary run, where $\beta^* = (\beta_1^*,\ \beta_{[-1]})$.

Step 2: Sample β_1^{*t} by the Random-Walk Metropolis algorithm as in Step 1 in the preliminary run.

Step 3: Sample $\beta_{[-1]}^t$ by the Multivariate Random-Walk Metropolis algorithm as in Step 2 in the preliminary run except that the variance-covariance matrix is updated using the variance-covariance matrix of the posterior sequence of $\beta_{[-1]}$ from the preliminary run.

Step 4: Sample σ from $IG(T, RQ(\beta^t, \theta))$.

After a burn-in period (B), calculate β_1 by (4.11)-(4.13) for SAV, AS, and indirect GARCH models, respectively, and we obtain the posterior mean for β from the second run.

4.4.4 Case Study (Continued)

We consider two scenarios for the second run, one without reparameterization and one with reparameterization to show that reparameterization improves the efficiency.

Scenario 1: Without Reparameterization

It took about 7.3 minutes to obtain a chain of length 60,000 for the four parameters, and the AR values are 0.1983 and 0.1497 for β_1 and $\beta_{[-1]}$, respectively. AR increases from 0.0322 in the preliminary run to 0.1497 in the second run.

Scenario 2: With Reparameterization

It took about 8.7 minutes to obtain a chain of length 60,000 for the four parameters. AR are 0.1921 and 0.6024 for β_1 and $\beta_{[-1]}$, respectively. AR increases from 0.0322 in the preliminary run to 0.6024 in the second run. Therefore, reparameterization can greatly improves the efficiency of the chain in the second run.

Summary statistics (mean, median, and standard deviation) are calculated for both scenarios to determine the burn-in period. According to Figure 4.3, all summary statistics are not stable for Scenario 1 but are stable after 20,000 iterations for Scenario 2. Therefore, we choose the burn-in period to be 20,000 for Scenario 2.

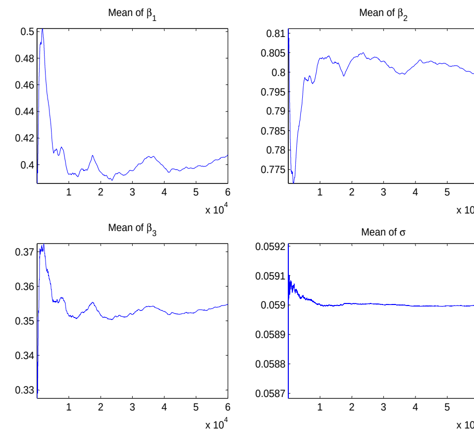
To test whether the convergence of the chain depends on the starting values, we perform the Gelman Rubin diagnostic (GRD) provided in Gelman and Rubin (1992). GRD is based on simulation of m independent chains with starting values drawn from the same starting distributions, each of length n after discarding the burn-in period. Denote the ith chain as $\{x_{it}\}_{t=1,\cdots,n}$, where $i = 1,\cdots,m$. Between-chain variability, B, and within-chain variability, W, are defined as

$$\begin{aligned}
\frac{B}{n} &= \text{the variance between the } m \text{ chain means } \bar{x}_{i\cdot} \\
&= \sum_{i=1}^{m}(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2/(m-1) \\
W &= \text{the average of the } m \text{ within-chain variances } s_i^2 \\
&= \sum_{i=1}^{m} s_i^2/m,
\end{aligned}$$

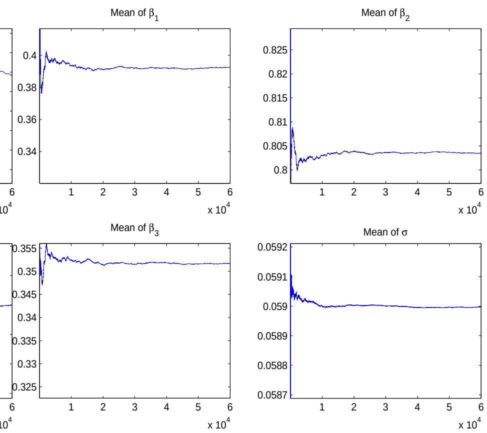
where $s_i^2 = \sum_{t=1}^{n}(x_{it} - \bar{x}_{i\cdot})^2/(n-1)$.

The total variability can be estimated by a weighted average of W and B,

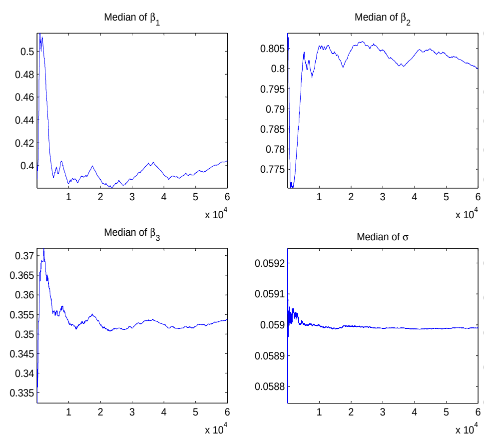
$$\hat{\sigma}_x^2 = \frac{n-1}{n}W + \frac{1}{n}B.$$



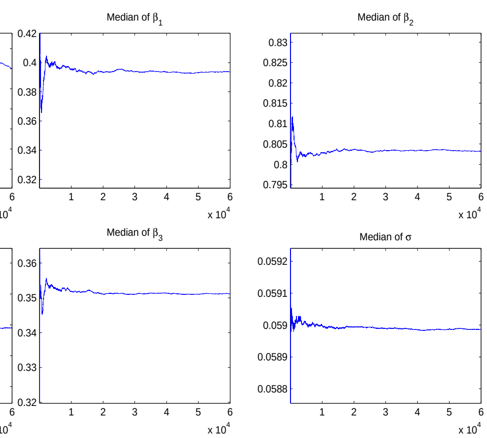
(a)



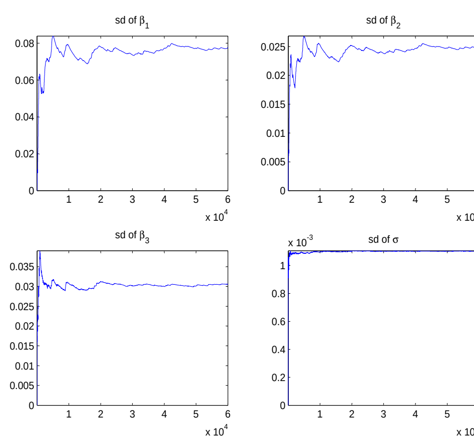
(a)



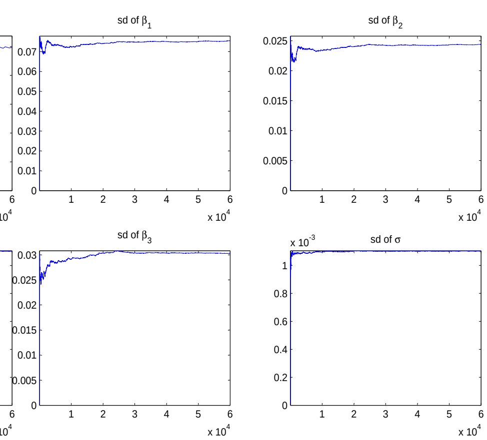
(b)



(b)



(c)



(c)

Figure 4.3: Summary statistics for each parameter in the second run: (a) mean . (b) median. (c) standard deviation. The left plot refers to without reparameterization, and the right plot refers to with reparameterization.

The potential scale reduction is estimated by

$$\hat{R} = \frac{\hat{\sigma}_x^2}{W} = \frac{n-1}{n} + \frac{1}{n}\frac{B}{W} > 1.$$

We expect $\hat{R}$ to be close to 1 when the chains converge.

After a preliminary run as discussed in Subsection 4.4.2, we obtain a variance-covariance matrix of $\beta_{[-1]}$ used in the proposal distribution for $\beta_{[-1]}$, and a VaR series ($\mathrm{VaR}_{t,\theta}(\hat{\beta})$, where $\hat{\beta}$ is the posterior mean) used for reparameterization. We generate 10 chains of length 60,000 using the updated variance-covariance matrix and the starting values, $(\beta_1^{*0}, \beta_2^0, \beta_3^0, \sigma^0)$, drawn from $U(0,1)$, respectively. We discard the burn-in period 20,000 for each chain. Therefore, we will have 10 independent chains of length 40,000 in the end. The estimated potential scale reduction, $\hat{R}$, are $1.0003, 1.0002, 1.0004$, and 1.000 for β_1, β_2, β_3, and σ, respectively. Because all $\hat{R}$ are less than 1.1, we see no evidence against convergence, and no evidence that the starting values are important.

We conclude that the two stage block algorithm proposed in Subsection 4.4.1 and 4.4.3 is more efficient compared to the element-wise MCMC implementation provided in Subsection 4.3.1. The AR values increase from 0.0518 and 0.1642 for β_2 and β_3, respectively, to 0.6024 for $\beta_{[-1]}$.

In addition, we perform the same sampling schemes except for Step 1, where β_1^{*t} is sampled from the neighborhood of the mode of the target distribution. Without reparameterization, it took about 9.3 minutes to get a chain of length 60,000 for the four parameters (three coefficients and one scale parameter) in the 1% SAV model. The AR values are 0.1521 and 0.5505 for β_1 and $\beta_{[-1]}$, respectively. With reparameterization, it took about 11.0 minutes to get a chain of length 60,000 for the four parameters (three coefficients and one scale parameter) in the 1% SAV model. The AR values are 0.1607 and 0.1126 for β_1 and $\beta_{[-1]}$, respectively. These results show that AR decreases over the Random-Walk Metropolis.

4.5 Empirical Investigations on Stable Period

To compare results from the Bayesian approach discussed in the previous section with the estimation method of Engle and Manganelli (2004), we use the same data analyzed by Engle and Manganelli (2004). They took a sample of 3,392 daily prices from April 7, 1986 to April 7, 1999 for three datasets, GM stock price, IBM stock price, and the S&P 500 index. The daily returns are computed as the difference in the log of successive prices. The first 2,892 (T) daily returns are treated as the training data or in-sample observations to estimate the model, and the last 500 (N) daily returns are treated as the testing data or out-of-sample observations to access the out-of-sample performance.

4.5.1 Results

According to the sampling schemes provided in Subsections 4.4.1 and 4.4.3, we generate a posterior sequence, $\{\beta^1, \cdots, \beta^M\}$, and collet the last J values for the Bayesian inference. Here, J is the length of the subsequence after excluding the burn-in period. According to the case study in the previous section, we choose $M = 60,000$, $J = 40,000$, $\nu_1 = \nu_2 = 0.05$, and small α_0 and s_0. Besides, we choose the posterior mean as the parameter estimate for β, and take the 0.025th/0.975th quantiles of the posterior subsequence as the lower bounds/upper bounds (LBs/UBs) of the credible intervals for β. For each vector in the posterior subsequence, β^j, we calculate VaR series ($\text{VaR}_{t,\theta}(\beta^j), t = 1, \cdots, T+N$) according to any of the four CAViaR models. For each VaR series, we further calculate the Hits series ($\text{Hits}_t(\beta^j)$) and the RQ value ($RQ(\beta^j, \theta)$). Hits in-sample sequence ($\text{Hits}_{IS}(\beta^j)$) is the average of the first T Hits series, and Hits out-of-sample sequence ($\text{Hits}_{OOS}(\beta^j)$) is the average of the last N Hits series, i.e., $\text{Hits}_{IS}(\beta^j) = \sum_1^T \text{Hits}_t(\beta^j)/T$, $\text{Hits}_{OOS}(\beta^j) = \sum_{T+1}^{T+N} \text{Hits}_t(\beta^j)/N$, and $\text{Hits}_t(\beta^j)$ is defined as $I(y_t < -\text{VaR}_{t,\theta}(\beta^j)) - \theta$. We use the posterior means of the Hits in-sample sequence and the Hits out-of-sample sequence as the estimates for Hits in-sample and Hits out-of-sample, and calculate the corresponding credible intervals and Monte Carlo standard errors (MCSEs) for the posterior mean. In addition, we use the posterior mean of RQ as the estimate for the RQ measure, and calculate the corresponding credible interval and MCSE for the posterior mean. The detail for calculating MCSE is based on Flegal et al. (2008). Suppose we have a posterior sequence $b^j, j = 1, \cdots, J$ for b, we use posterior mean $\bar{b} = \sum_1^J b^j/J$ to estimate b. According to the CLT,

$$\sqrt{J}(\bar{b} - E(b|\text{data})) \to N(0, \sigma_b^2) \text{ in distribution as } J \to \infty$$

We break the sequence into blocks of equal size, and calculate batch means estimate of b. Suppose that c_J is the batch size, a_J is the number of batches, $a_J c_J = J$, and $\bar{b}_k$ is the batch mean for each block, i.e.,

$$\bar{b}_k = \frac{1}{c_J} \sum_{j=(k-1)c_J+1}^{kc_J} b^j, \ k = 1, \cdots, a_J.$$

The batche means estimate of σ_b^2 is

$$\hat{\sigma}_b^2 = \frac{c_J}{a_J - 1} \sum_{k=1}^{a_J} (\bar{b}_k - \bar{b})^2.$$

Finally, the MCSE for $\bar{b}$ is $\hat{\sigma}_b/\sqrt{J}$. Flegal et al. (2008) suggests that $c_J = [\sqrt{J}]$, and $a_J = [J/c_J]$ is a convenient choice that works well in applications. Therefore, we use MCSE to access the accuracy of the point estimate, $\bar{b}$. We have several interesting findings based on Table 4.1 to Table 4.3. Compared to the confidence intervals constructed by Engle and Manganelli (2004), we find that the length of the intervals from our Bayesian approach is much shorter. Besides, credible intervals and confidence intervals for β are overlapping except for the intercept β_1 in the SAV model on the S&P 500 index. When we use Engle and Manganelli's approach, we will obtain a higher estimate for β_1 in this case. In addition, Engle and Manganelli's method of estimation did not find the optimal quantile estimates that minimize the objective function, RQ, because smaller RQ was attained in the SAV model for the S&P 500 data by the Bayesian approach. The smallest RQ appeared in the AS model for both approaches. Therefore, the AS model has the largest Pseudo-R^2 according to equation (4.2). Importantly, the performance of CAViaR models is better using the Bayesian approach than Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria. The posterior estimates for Hits in-sample and Hits out-of-sample are very close to the theoretical value, 0, but not for Engle and Manganelli's approach. The smallest Hits out-of-sample measure tends to appear in the Adaptive model. The posterior distributions of Hits in-sample and Hits out-of-sample are provided in Figure 4.4-Figure 4.7. In the Hits in-sample plots, Hits are centered around the theoretical value, 0, for the SAV, AS, and indirect GRACH models, but not so for the Adaptive model. Therefore, the Adaptive model does not fit the data well. In the Hits out-of-sample plots, although the Hits are not centered around 0, the mean is closer to zero than Engle and Manganelli's method. Take 1% GM Hits out-of-sample using the SAV model for example, the range is $[0, 0.004]$ for the Bayesian approach, as compared to 0.4 for Engle and Manganelli's method. Besides, we plot the estimated 1% VaRs for the first 10 days in the out-of-sample period with credible intervals for GM. Figure 4.8 shows similar patterns but different scales of the estimated VaR from the Bayesian approach and Engle and Manganelli's approach. The estimated VaRs are larger from the Bayesian approach compared to Engle and Manganelli's approach except for the Adaptive model for these 10 days. We found that the fluctuation from the Adaptive model is very mild due to the unit coefficient on the lagged VaR. Returns in the third and fourth days are 1.4907 and -3.4512, respectively. Therefore, we have the highest VaR in the fifth day from all models.

Table 4.1: Comparison of estimates and relevant statistics for the four CAViaR models between (a) Bayesian and (b) Engle and Manganelli's approaches using GM data from April 7, 1986 to April 7, 1999. LB/UB refers to lower/upper bound of credible interval form the Bayesian approach or lower/upper bound of confidence interval from Engle and Manganelli's approach. MCSE refers to Monte Carlo standard error computed based on Flegal et al. (2008)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1% VaR								
$\boldsymbol{\beta}_1$	**0.3913**	**0.4511**	**0.4076**	0.3734	**1.4794**	1.4959	**0.3092**	**0.2968**
LB	0.2527	0.0536	0.3205	-0.1005	0.9060	-0.3175	0.2731	0.0794
UB	0.5394	0.8486	0.5239	0.8473	2.1252	3.3093	0.3721	0.5142
$\boldsymbol{\beta}_2$	**0.8042**	**0.8263**	**0.7907**	**0.7995**	**0.7799**	**0.7804**		
LB	0.7568	0.6644	0.7536	0.6292	0.7266	0.6648		
UB	0.8493	0.9882	0.8213	0.9698	0.8283	0.8960		
$\boldsymbol{\beta}_3$	**0.3503**	**0.3305**	**0.2844**	**0.2779**	**0.9519**	0.9356		
LB	0.2913	0.0002	0.2360	0.0039	0.7612	-1.5377		
UB	0.4121	0.6608	0.3408	0.5519	1.1470	3.4089		
$\boldsymbol{\beta}_4$			**0.4638**	**0.4569**				
LB			0.3876	0.1066				
UB			0.5390	0.8072				
$\boldsymbol{\sigma}$	0.0590		0.0586		0.0592		0.0621	
LB	0.0569		0.0565		0.0571		0.0599	
UB	0.0612		0.0608		0.0614		0.0644	
RQ	**170.57**	172.04	169.38	169.22	171.11	170.99	179.66	179.61
MCSE	0.0011		0.0025		0.0016		0.0014	
LB	170.50		169.25		171.00		179.61	
UB	170.76		169.62		171.31		179.81	
Hits in-sample (%)	**0.0000**	0.0028	**0.0000**	-0.0318	**0.0000**	0.0028	**0.0002**	-0.0318
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	-0.0010		-0.0014		-0.0010		0.0000	
UB	0.0011		0.0014		0.0011		0.0011	
Hits out-of-sample (%)	**0.0014**	0.4000	**0.0028**	0.4000	**0.0020**	0.2000	**0.0079**	0.8000
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	0.0000		0.0000		0.0020		0.0080	
UB	0.0040		0.0040		0.0020		0.0080	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

Table 4.1 (cont.)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
5% VaR								
$\boldsymbol{\beta}_1$	**0.2330**	**0.1812**	**0.0773**	**0.0760**	**0.4169**	**0.3336**	**0.2856**	**0.2871**
LB	0.1106	0.0179	0.0473	0.0272	0.2312	0.1300	0.2142	0.1879
UB	0.4191	0.3445	0.1268	0.1248	0.9485	0.5372	0.3473	0.3863
$\boldsymbol{\beta}_2$	**0.8510**	**0.8953**	**0.9333**	**0.9326**	**0.8902**	**0.9042**		
LB	0.7677	0.8245	0.9070	0.8946	0.7934	0.8779		
UB	0.9099	0.9661	0.9513	0.9706	0.9328	0.9305		
$\boldsymbol{\beta}_3$	**0.1287**	**0.1133**	**0.0380**	0.0398	**0.1243**	0.1220		
LB	0.0949	0.0894	0.0149	-0.0233	0.0704	-0.1032		
UB	0.1699	0.1372	0.0635	0.1029	0.1906	0.3472		
$\boldsymbol{\beta}_4$			**0.1183**	**0.1218**				
LB			0.0915	0.0424				
UB			0.1396	0.2012				
$\boldsymbol{\sigma}$	**0.1908**		**0.1898**		**0.1911**		**0.1916**	
LB	0.1841		0.1830		0.1843		0.1846	
UB	0.1979		0.1968		0.1982		0.1986	
RQ	551.72	550.83	548.77	548.31	552.55	552.12	553.91	553.79
MCSE	0.0095		0.0083		0.0092		0.0031	
LB	551.36		548.37		552.17		553.79	
UB	552.39		549.55		553.41		554.31	
Hits in-sample (%)	**-0.0001**	-0.0207	**-0.0002**	-0.0899	**0.0000**	-0.0207	**-0.0014**	-0.0899
MCSE	0.0000		0.0001		0.0000		0.0000	
LB	-0.0033		-0.0040		-0.0040		-0.0019	
UB	0.0036		0.0039		0.0036		-0.0009	
Hits out-of-sample (%)	**-0.0054**	-0.2000	0.0026	0.0000	**-0.0033**	-0.4000	**0.0098**	1.0000
MCSE	0.0000		0.0001		0.0000		0.0000	
LB	-0.0080		-0.0020		-0.0080		0.0080	
UB	0.0000		0.0080		0.0000		0.0120	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

Table 4.2: Comparison of estimates and relevant statistics for the four CAViaR models between (a) Bayesian and (b) Engle and Manganelli's approaches using IBM data from April 7, 1986 to April 7, 1999. LB/UB refers to lower/upper bound of credible interval form the Bayesian approach or lower/upper bound of confidence interval from Engle and Manganelli's approach. MCSE refers to Monte Carlo standard error computed based on Flegal et al. (2008)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1% **VaR**								
$\boldsymbol{\beta}_1$	**0.1705**	0.1261	**0.0685**	0.0558	**1.5969**	1.3289	**0.1729**	**0.1626**
LB	0.0583	-0.0560	0.0428	-0.0500	1.0095	-2.4907	0.1293	0.0183
UB	0.2899	0.3082	0.1000	0.1616	2.3661	5.1485	0.2273	0.3069
$\boldsymbol{\beta}_2$	**0.9107**	**0.9476**	**0.9433**	**0.9423**	**0.8568**	**0.8740**		
LB	0.8591	0.8494	0.9291	0.8939	0.8090	0.6519		
UB	0.9585	1.0458	0.9558	0.9907	0.8939	1.0961		
$\boldsymbol{\beta}_3$	**0.1742**	0.1134	**0.0395**	0.0499	**0.3517**	**0.3374**		
LB	0.0878	-0.1189	0.0041	-0.0604	0.3135	0.1506		
UB	0.2711	0.3457	0.0766	0.1602	0.4108	0.5242		
$\boldsymbol{\beta}_4$			**0.2336**	**0.2512**				
LB			0.1929	0.0850				
UB			0.2693	0.4174				
$\boldsymbol{\sigma}$	**0.0632**		**0.0621**		**0.0635**		**0.0665**	
LB	0.0610		0.0599		0.0612		0.0641	
UB	0.0655		0.0644		0.0658		0.0690	
RQ	182.84	182.32	179.53	179.40	183.56	183.43	192.25	192.20
MCSE	0.0036		0.0018		0.0017		0.0011	
LB	182.68		179.43		183.45		192.20	
UB	183.09		179.75		183.81		192.40	
Hits in-sample (%)	**0.0000**	-0.0318	**0.0000**	0.0373	**0.0000**	0.0028	**0.0023**	0.2448
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	-0.0014		-0.0010		-0.0014		0.0021	
UB	0.0011		0.0011		0.0011		0.0028	
Hits out-of-sample (%)	**0.0051**	0.6000	**0.0062**	0.6000	**0.0066**	0.6000	**0.0051**	0.6000
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	0.0040		0.0040		0.0040		0.0040	
UB	0.0060		0.0080		0.0080		0.0080	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

Table 4.2 (cont.)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
5% VaR								
$\boldsymbol{\beta_1}$	**0.0532**	0.1191	**0.0999**	**0.0953**	**0.6240**	**0.5387**	**0.3976**	**0.3969**
LB	0.0177	-0.0453	0.0598	-0.0090	0.2385	0.2312	0.3497	0.2377
UB	0.1143	0.2835	0.1528	0.1996	0.9022	0.8462	0.4459	0.5561
$\boldsymbol{\beta_2}$	**0.9118**	**0.9053**	**0.8895**	**0.8892**	**0.8024**	**0.8259**		
LB	0.8726	0.8073	0.8553	0.8137	0.7350	0.7683		
UB	0.9377	1.0033	0.9163	0.9647	0.9016	0.8835		
$\boldsymbol{\beta_3}$	**0.1363**	**0.1481**	**0.0616**	**0.0617**	**0.1767**	0.1591		
LB	0.1072	0.0799	0.0392	0.0084	0.1181	-0.0667		
UB	0.1675	0.2163	0.0870	0.1150	0.2351	0.3849		
$\boldsymbol{\beta_4}$			**0.2092**	**0.2187**				
LB			0.1713	0.1276				
UB			0.2524	0.3098				
$\boldsymbol{\sigma}$	**0.1805**		**0.1785**		**0.1817**		**0.1825**	
LB	0.1740		0.1721		0.1751		0.1760	
UB	0.1872		0.1851		0.1884		0.1893	
RQ	**521.83**	522.43	516.01	515.58	525.10	524.79	527.80	527.72
MCSE	0.0067		0.0067		0.0042		0.0026	
LB	521.54		515.65		524.82		527.72	
UB	522.40		516.76		525.69		528.19	
Hits in-sample (%)	**-0.0002**	0.0138	**-0.0001**	-0.0207	**0.0001**	0.0484	**-0.0015**	-0.1591
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	-0.0040		-0.0044		-0.0040		-0.0019	
UB	0.0043		0.0046		0.0050		-0.0006	
Hits out-of-sample (%)	**0.0068**	1.0000	**0.0252**	2.4000	**0.0231**	2.4000	**0.0001**	0.0000
MCSE	0.0001		0.0000		0.0001		0.0000	
LB	-0.0020		0.0200		0.0160		0.0000	
UB	0.0140		0.0280		0.0300		0.0020	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

Table 4.3: Comparison of estimates and relevant statistics for the four CAViaR models between (a) Bayesian and (b) Engle and Manganelli's approaches using S&P 500 data from April 7, 1986 to April 7, 1999. LB/UB refers to lower/upper bound of credible interval form the Bayesian approach or lower/upper bound of confidence interval from Engle and Manganelli's approach. MCSE refers to Monte Carlo standard error computed based on Flegal et al. (2008)

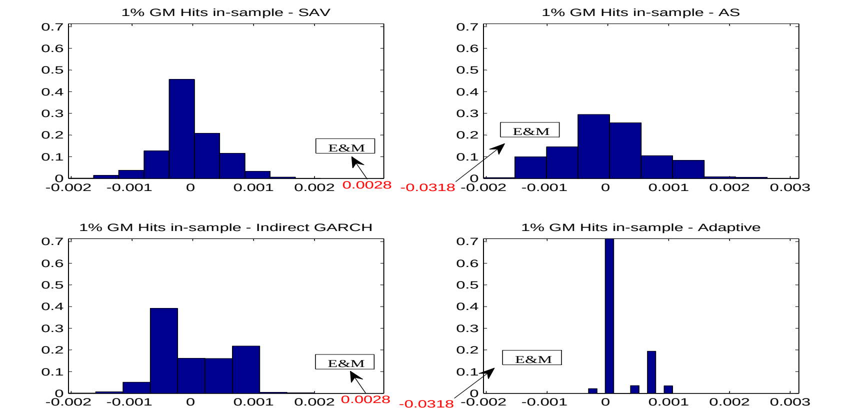
	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1% VaR								
$\boldsymbol{\beta}_1$	**0.0323**	**0.2039**	**0.1453**	**0.1476**	**0.2240**	**0.2328**	**0.5594**	**0.5562**
LB	0.0022	0.0855	0.1263	0.0582	0.1703	-0.0006	0.5203	0.3308
UB	0.0784	0.3223	0.1653	0.2370	0.2816	0.4662	0.6064	0.7816
$\boldsymbol{\beta}_2$	**0.9332**	**0.8732**	**0.8745**	**0.8729**	**0.8480**	**0.8350**		
LB	0.8982	0.7738	0.8622	0.8137	0.8318	0.7909		
UB	0.9585	0.9726	0.8892	0.9321	0.8673	0.8791		
$\boldsymbol{\beta}_3$	**0.1939**	0.3819	**-0.0077**	-0.0139	**0.9229**	1.0582		
LB	0.1473	-0.1614	-0.0432	-0.2389	0.7223	-1.0945		
UB	0.2632	0.9252	0.0342	0.2111	1.1329	3.2109		
$\boldsymbol{\beta}_4$			**0.4820**	**0.4969**				
LB			0.4108	0.2339				
UB			0.5251	0.7599				
$\boldsymbol{\sigma}$	**0.0373**		**0.0366**		**0.0375**		**0.0406**	
LB	0.0360		0.0353		0.0361		0.0392	
UB	0.0387		0.0380		0.0389		0.0421	
RQ	**107.93**	109.68	105.95	105.82	108.40	108.34	117.44	117.42
MCSE	0.0017		0.0018		0.0010		0.0007	
LB	107.84		105.85		108.35		117.42	
UB	108.06		106.11		108.51		117.53	
Hits in-sample (%)	**0.0000**	0.0028	**-0.0001**	-0.0318	**0.0000**	0.0028	**-0.0007**	-0.0664
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	-0.0010		-0.0014		-0.0010		-0.0007	
UB	0.0011		0.0014		0.0011		-0.0007	
Hits out-of-sample (%)	**0.0022**	0.8000	**0.0068**	0.6000	**0.0077**	0.8000	**0.0020**	0.2000
MCSE	0.0000		0.0000		0.0000		0.0000	
LB	0.0000		0.0040		0.0040		0.0020	
UB	0.0080		0.0100		0.0080		0.0020	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

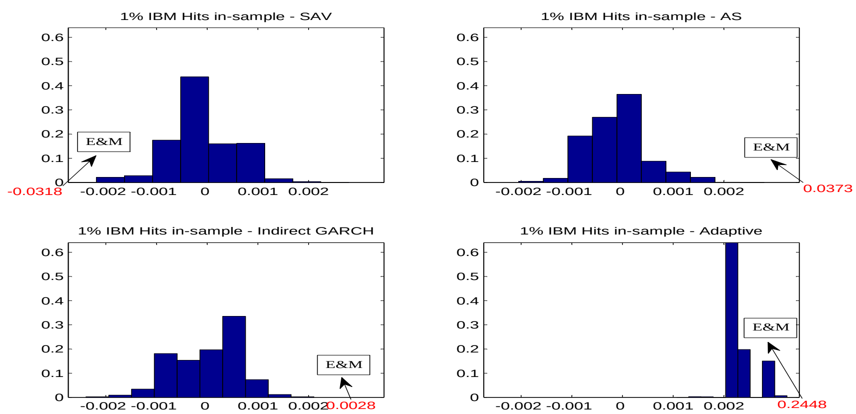
Table 4.3 (cont.)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
5% VaR								
$\boldsymbol{\beta}_1$	**0.0095**	**0.0511**	**0.0619**	**0.0378**	**0.0302**	**0.0262**	**0.3197**	**0.3700**
LB	0.0031	0.0348	0.0321	0.0113	0.0185	0.0066	0.2067	0.2197
UB	0.0166	0.0674	0.0951	0.0643	0.0445	0.0458	0.4063	0.5203
$\boldsymbol{\beta}_2$	**0.9501**	**0.9369**	**0.8781**	**0.9025**	**0.9261**	**0.9287**		
LB	0.9332	0.8930	0.8381	0.8743	0.9132	0.9167		
UB	0.9664	0.9808	0.9124	0.9307	0.9383	0.9407		
$\boldsymbol{\beta}_3$	**0.0912**	**0.1341**	**0.0266**	**0.0377**	**0.1377**	0.1407		
LB	0.0620	0.0328	-0.0026	-0.0062	0.1115	-1.0741		
UB	0.1238	0.2354	0.0569	0.0816	0.1579	1.3555		
$\boldsymbol{\beta}_4$			**0.3257**	**0.2871**				
LB			0.2571	0.2365				
UB			0.3993	0.3377				
$\boldsymbol{\sigma}$	**0.1061**		**0.1041**		**0.1059**		**0.1079**	
LB	0.1023		0.1004		0.1021		0.1040	
UB	0.1100		0.1080		0.1098		0.1120	
RQ	**306.67**	306.68	301.07	300.82	306.15	305.93	312.14	312.06
MCSE	0.0034		0.0038		0.0034		0.0024	
LB	306.53		300.87		305.96		312.06	
UB	307.01		301.46		306.49		312.39	
Hits in-sample (%)	**-0.0001**	0.0484	**-0.0001**	0.0138	**0.0004**	0.0138	**-0.0029**	-0.2628
MCSE	0.0001		0.0000		0.0000		0.0000	
LB	-0.0037		-0.0040		-0.0033		-0.0040	
UB	0.0043		0.0036		0.0039		-0.0019	
Hits out-of-sample (%)	**0.0045**	0.6000	**0.0241**	1.4000	**0.0099**	0.8000	**-0.0026**	-0.4000
MCSE	0.0000		0.0001		0.0001		0.0000	
LB	0.0000		0.0140		0.0060		-0.0060	
UB	0.0080		0.0360		0.0160		0.0000	

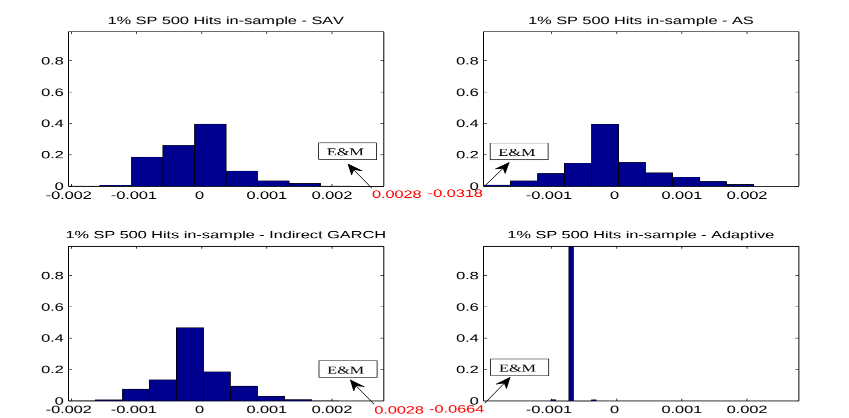
Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.



(a) 1% GM Hits in-sample

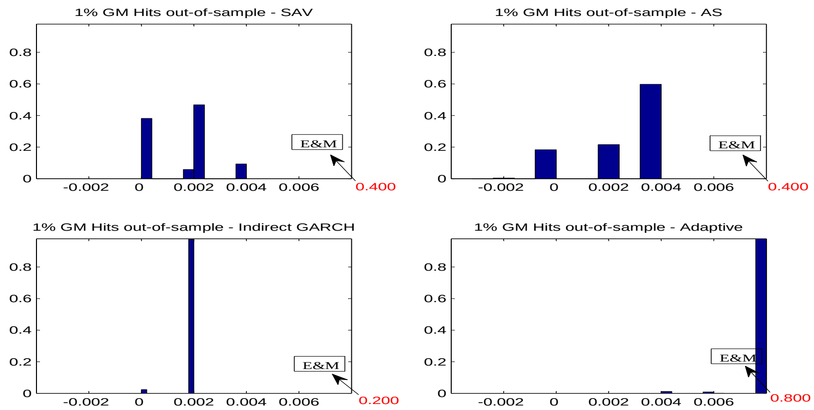


(b)1% IBM Hits in-sample

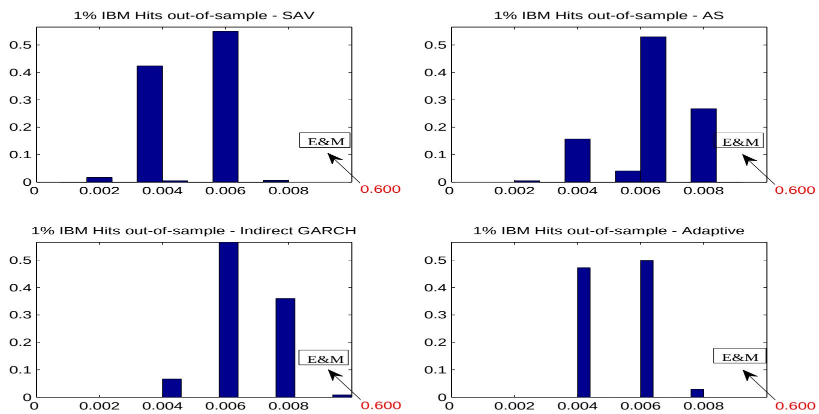


(c)1% S&P 500 Hits in-sample

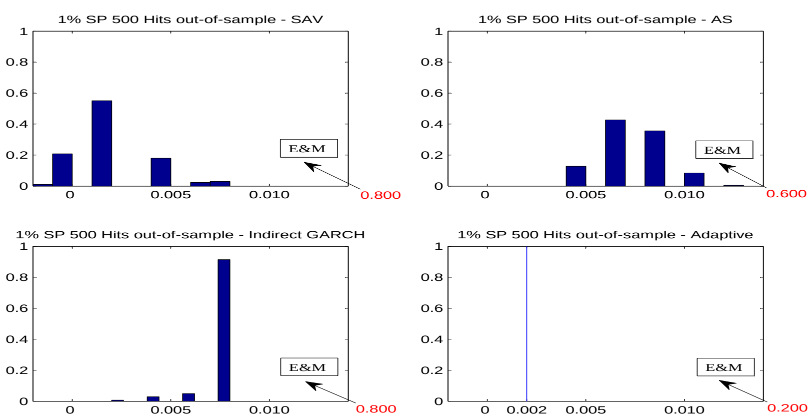
Figure 4.4: Relative Frequency of 1% Hits in-sample for each of the three datasets. Hits in-sample are centered around the theoretical value, 0, compared to the Engle and Manganelli's point estimates that are not in the scale of the box.



(a)1% GM Hits out-of-sample

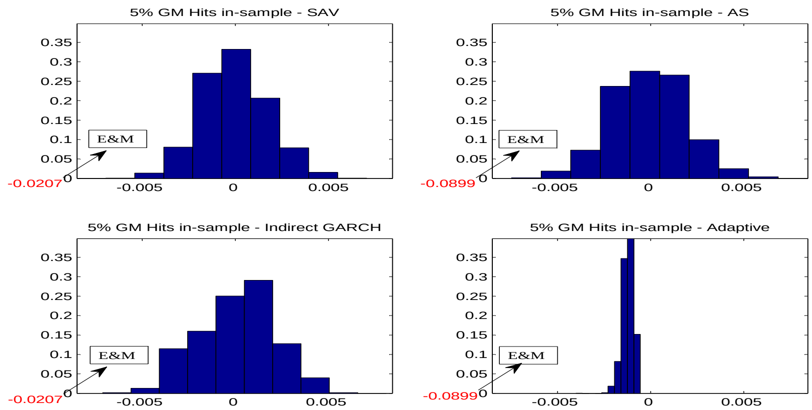


(b)1% IBM Hits out-of-sample

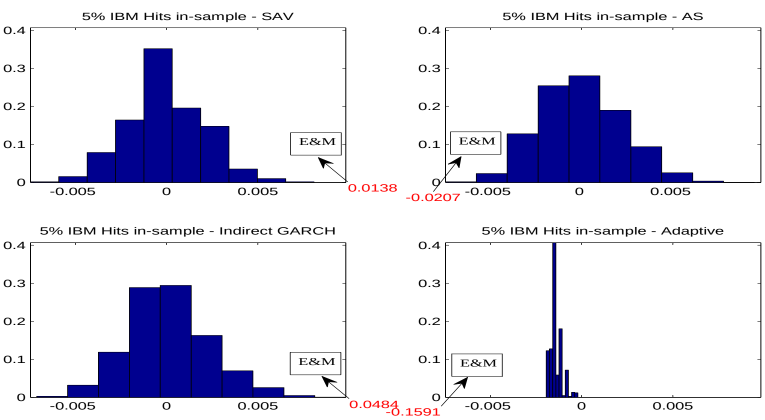


(c)1% S&P 500 Hits out-of-sample

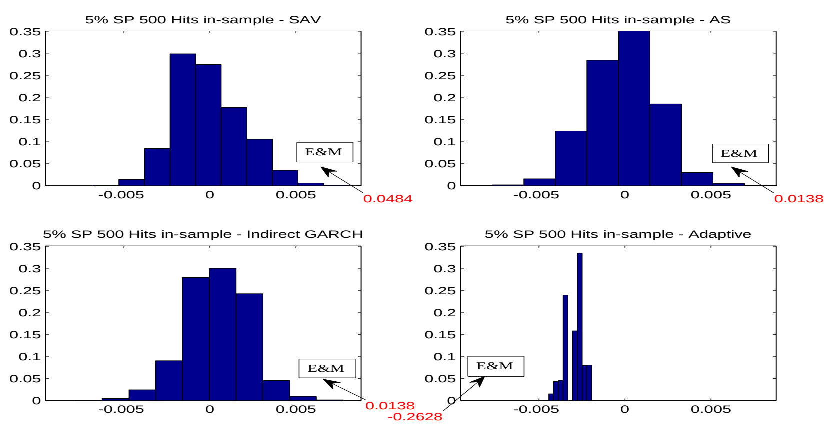
Figure 4.5: Relative Frequency of 1% Hits out-of-sample for each of the three datasets. Although almost all Hits out-of-sample are not centered around 0, the means are closer to zero compared to Engle and Manganelli's point estimates that are not in the scale of the box.



(a) 5% GM Hits in-sample

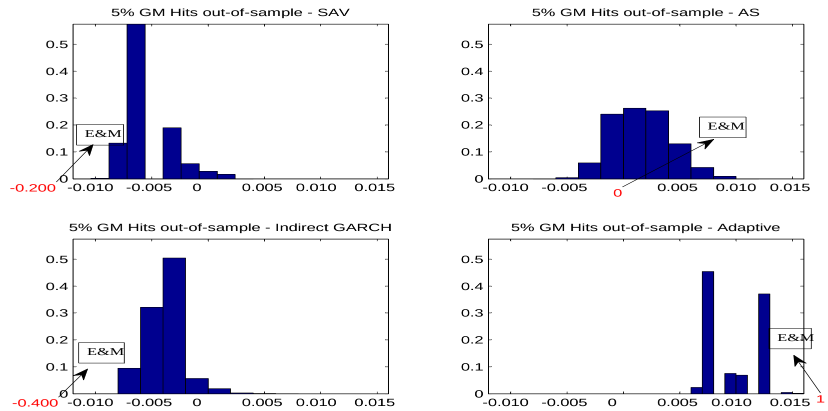


(b)5% IBM Hits in-sample

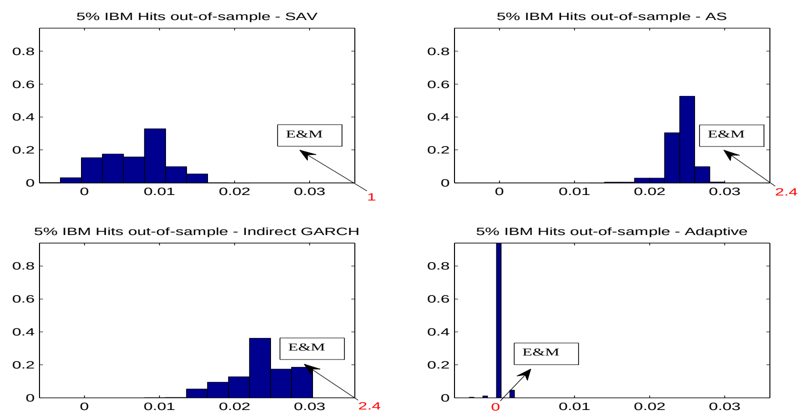


(c)5% S&P 500 Hits in-sample

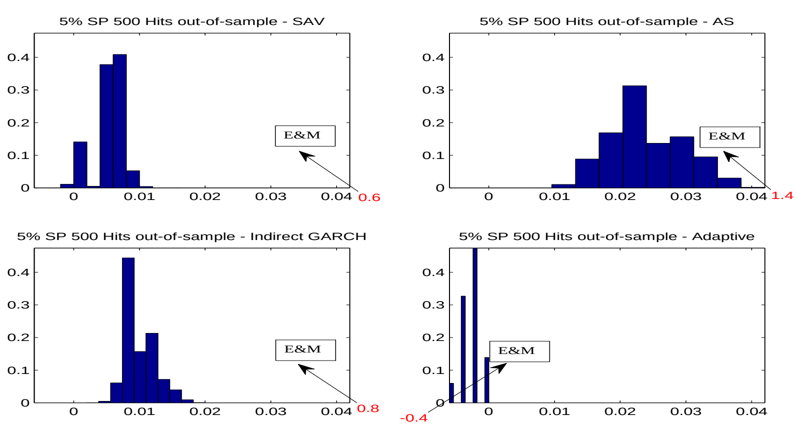
Figure 4.6: Relative Frequency of 5% Hits in-sample for each of the three datasets. Hits in-sample are centered around the theoretical value, 0, compared to the Engle and Manganelli's point estimates that are not in the scale of the box.



(a)5% GM Hits out-of-sample

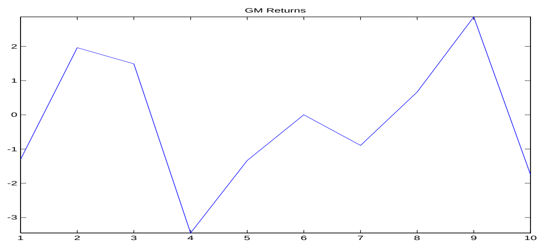


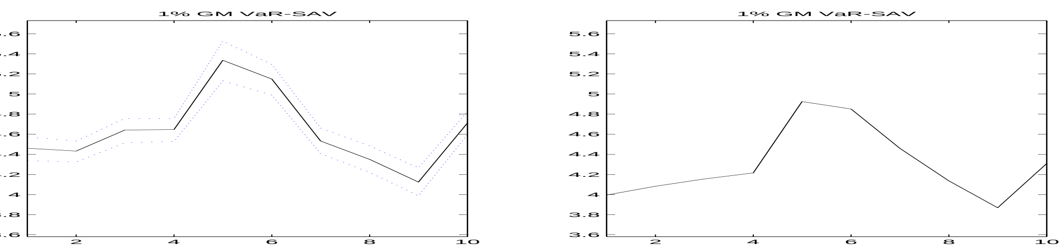
(b)5% IBM Hits out-of-sample



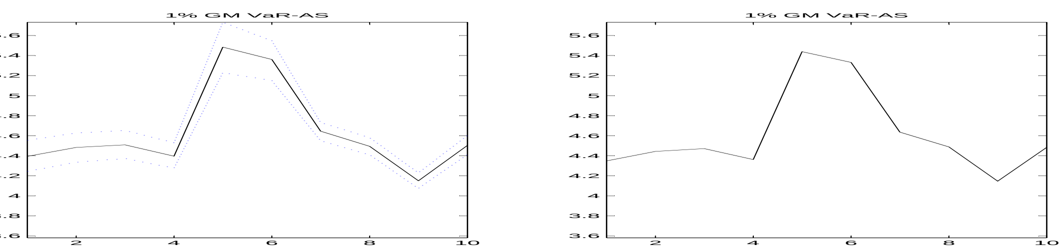
(c)5% S&P 500 Hits out-of-sample

Figure 4.7: Relative Frequency of 5% Hits out-of-sample for each of the three datasets. Although almost all Hits out-of-sample are not centered around 0, the means are closer to zero compared to Engle and Manganelli's point estimates that are not in the scale of the box.

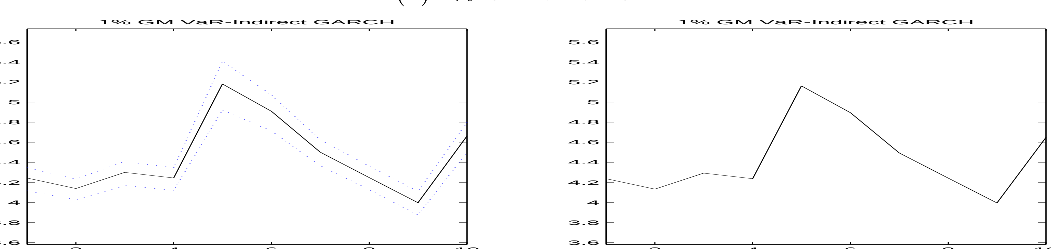




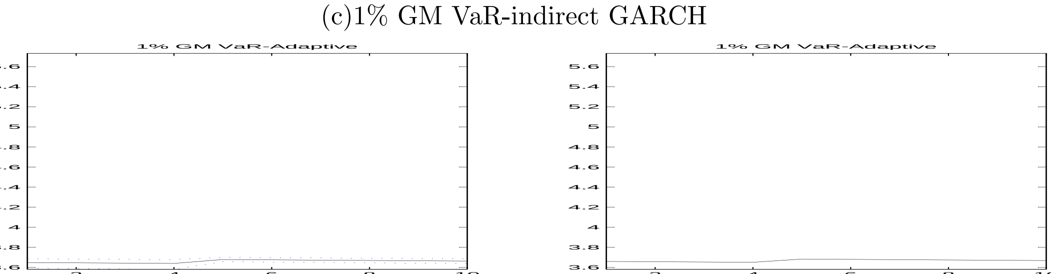
(a) 1% GM VaR-SAV



(b) 1% GM VaR-AS



(c)1% GM VaR-indirect GARCH



(d)1% GM VaR-Adaptive

Figure 4.8: Returns and 1% estimated VaR plots for GM for the first 10 days in the out-of-sample period. The left plot with 95 % pointwise credible bands is based on the Bayesian approach, and the right plot is based on Engle and Manganelli's approach.

4.6 Empirical Investigations on Global Financial Crisis

In this section, we investigate the performance of our method in a volatile period. We use a sample of 3,525 daily indices from January 2, 1996 to January 2, 2010 for the S&P 500 index. Daily returns are computed as arithmetic return (the ratio of the difference between two successive indices to the first index). Figure 4.9 shows the indices and returns over this period including global financial crisis. We choose September 15, 2008 as the cutoff to split the data into two sets, training period and testing period. The first 3,197 (T) daily returns are treated as the training data or in-sample observations to estimate the model, and the last 328 (N) daily returns are treated as the testing data or out-of-sample observations to assess the out-of-sample performance.

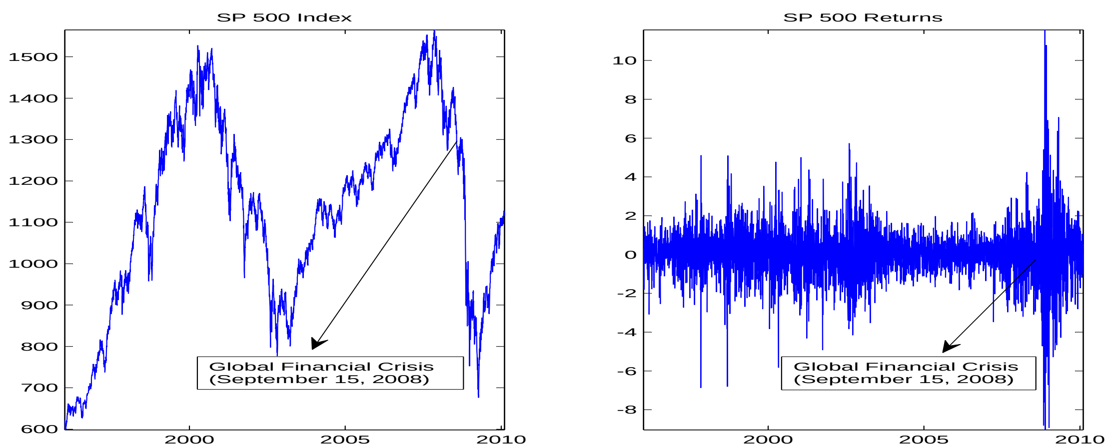


Figure 4.9: S&P 500 index from January 2, 1996 to January 2, 2010. September 15, 2008 is chosen as the cutoff to split the data into training period and testing period.

4.6.1 Results

We choose $M = 60,000$, $J = 40,000$, $\nu_1 = \nu_2 = 0.005$ for the MCMC algorithm described in Section 4.4. We have several interesting findings based on Table 4.4. Compared to the confidence intervals from Engle and Manganelli's approach, we find that the length of the intervals from the Bayesian approach is much shorter. Besides, credible intervals and confidence intervals for β are overlapping. In addition, Engle and Manganelli's method of estimation did not find the optimal quantile estimates that minimize the objective function, RQ, because smaller RQ was attained on each CAViaR model by the Bayesian approach. The smallest RQ appeared in the AS model for both approaches. Therefore, the AS model has the largest Pseudo-R^2. Importantly, the performance of CAViaR models is better using the Bayesian approach than Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria. The posterior estimates for Hits in-sample and Hits out-of-sample are very close to the theoretical value, 0, but not for Engle and Manganelli's approach. The smallest Hits out-of-sample measure tends to appear in the Adaptive model. The posterior distributions of Hits in-sample and Hits out-of-sample are provided in Figure 4.10. In the Hits in-sample plots, Hits are centered around the theoretical value, 0, for the SAV, AS, and indirect GRACH models, but not so for the Adaptive model. Therefore, the Adaptive model does not fit the data well. In the Hits out-of-sample plots, although the Hits are not centered around 0, the mean is closer to zero than Engle and Manganelli's method. Take 1% Hits out-of-sample using the SAV model for example, the range is $[0.017, 0.034]$ for the Bayesian approach, as compared to 1.134 for Engle and Manganelli's method. Besides, we plot estimated 1% VaRs for the 30 days before and after cutoff (September 15, 2008) with credible intervals and confidence intervals for both approaches. Take the indirect GARCH as an example, we describe how to obtain the confidence interval for $\mathrm{VaR}_{t,\theta}(\beta)$. According to Theorems 2 and 3 of Engle and Manganelli (2004), we have

$$\sqrt{T} A_T^{-1/2} D_T(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, I),$$

where

$$\begin{aligned} A_T &\equiv E\left[T^{-1}\theta(1-\theta)\sum_{t=1}^{T} \nabla'\mathrm{VaR}_{t,\theta}(\beta)\nabla\mathrm{VaR}_{t,\theta}(\beta)\right], \\ D_T &\equiv E\left[T^{-1}\sum_{t=1}^{T} f_t \nabla'\mathrm{VaR}_{t,\theta}(\beta)\nabla\mathrm{VaR}_{t,\theta}(\beta)\right], \end{aligned}$$

f_t is the conditional density function of $\epsilon_{t,\theta}$ evaluated at 0, and $\nabla\mathrm{VaR}_{t,\theta}(\beta)$ is the gradient of $\mathrm{VaR}_{t,\theta}(\beta)$. Therefore,

$$Var(\hat{\beta}) = T^{-1} D_T^{-1} A_T D_T^{-1}.$$

Under some regularity conditions, the consistent estimators for A_T and D_T are

$$\begin{aligned} \hat{A}_T &= T^{-1}\theta(1-\theta)\sum_{t=1}^{T} \nabla'\mathrm{VaR}_{t,\theta}(\hat{\beta})\nabla\mathrm{VaR}_{t,\theta}(\hat{\beta}), \\ \hat{D}_T &= (2T\hat{c}_T)^{-1}\sum_{t=1}^{T} I(|y_t + \mathrm{VaR}_{t,\theta}(\hat{\beta})| < \hat{c}_T)\nabla'\mathrm{VaR}_{t,\theta}(\hat{\beta})\nabla\mathrm{VaR}_{t,\theta}(\hat{\beta}), \end{aligned}$$

where $\hat{c}_T$ is a bandwidth satisfies $c_T = o(1)$, $c_T^{-1} = o(T^{1/2})$, and $\hat{c}_T/c_T \to 0$ in probability. The indirect GRACH model can be written as

$$\mathrm{VaR}^2_{t,\theta}(\beta) = B_3\beta,$$

where $B_3 = (1,\ \mathrm{VaR}_{t-1,\theta}(\beta)^2,\ y^2_{t-1})$, and $\beta = (\beta_1,\ \beta_2,\ \beta_3)^T$. We have

$$Var(\hat{\mathrm{VaR}}^2_{t,\theta}(\hat{\beta})) = \hat{B}_3 Var(\hat{\beta})\hat{B}_3^T,$$

where $\hat{B}_3 = (1,\ \hat{\mathrm{VaR}}_{t-1,\theta}(\hat{\beta}),\ y^2_{t-1})$. According to the Delta method, we have

$$Var(\hat{\mathrm{VaR}}_{t,\theta}(\hat{\beta})) = \frac{1}{4\hat{\mathrm{VaR}}^2_{t,\theta}(\hat{\beta})}\hat{B}_3 Var(\hat{\beta})\hat{B}_3^T.$$

Therefore, $(1-\theta)$ confidence interval for $\mathrm{VaR}_{t,\theta}(\beta)$ is

$$\hat{\mathrm{VaR}}_{t,\theta}(\hat{\beta}) \pm Z_{\theta/2}\sqrt{Var(\hat{\mathrm{VaR}}_{t,\theta}(\hat{\beta}))},$$

where $Z_{\theta/2}$ denotes the $(1-\theta/2)$-th quantile of the standard normal distribution. Similarly, we can obtain confidence intervals for other CAViaR models. Figure 4.11 shows similar patterns of the estimated VaR, but different widths of CI from the Bayesian approach and Engle and Manganelli's approach. We found that the fluctuation of VaR from the Adaptive model is very mild due to the unit coefficient on the lagged VaR. Before the cutoff, we have small VaRs because of the stable returns. However, we have large VaRs after the cutoff because of the volatile returns. The confidence intervals from Engle and Manganelli's approach are much wider than the Bayesian approach, especially for the indirect GARCH model.

Table 4.4: Comparison of estimates and relevant statistics for the four CAViaR models between (a) Bayesian and (b) Engle and Manganelli's approaches using S&P 500 data from January 2, 1996 to January 2, 2010. LB/UB refers to lower/upper bound of credible interval form the Bayesian approach or lower/upper bound of confidence interval from Engle and Manganelli's approach. MCSE refers to Monte Carlo standard error computed based on Flegal et al. (2008)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1% VaR								
$\boldsymbol{\beta}_1$	**0.0568**	0.0462	**0.1338**	**0.1024**	**0.1167**	0.1234	**0.7619**	**0.6508**
LB	0.0439	-0.0085	0.1144	0.0309	0.0728	-0.0011	0.6400	0.3417
UB	0.0721	0.1009	0.1653	0.1739	0.1596	0.2479	0.8172	0.9599
$\boldsymbol{\beta}_2$	**0.9329**	**0.9731**	**0.8979**	**0.9189**	**0.9351**	**0.9413**		
LB	0.9203	0.9351	0.8758	0.8485	0.9233	0.9237		
UB	0.9457	1.0111	0.9131	0.9893	0.9505	0.9589		
$\boldsymbol{\beta}_3$	**0.1391**	**0.0641**	**-0.0684**	-0.0541	**0.2626**	0.2434		
LB	0.1101	0.0010	-0.0940	-0.2160	0.1976	-0.5253		
UB	0.1641	0.1272	-0.0436	0.1078	0.3110	1.0121		
$\boldsymbol{\beta}_4$			**0.3841**	0.2339				
LB			0.3249	-0.0209				
UB			0.4652	0.6635				
$\boldsymbol{\sigma}$	**0.0342**		**0.0331**		**0.0342**		**0.0354**	
LB	0.0329		0.0319		0.0329		0.0341	
UB	0.0354		0.0343		0.0354		0.0367	
RQ	**98.75**	110.87	**95.59**	107.05	**98.79**	110.22	**102.31**	112.18
MCSE	0.0012		0.0015		0.0006		0.0035	
LB	98.68		95.51		98.74		102.29	
UB	98.87		95.73		98.89		102.40	
Hits in-sample (%)	**0**	0.0009	**0**	0.0322	**0.0001**	0.0322	**-0.0009**	0.0948
MCSE	0		0		0		0	
LB	-0.0014		-0.0014		-0.0010		-0.0010	
UB	0.0011		0.0014		0.0018		0	
Hits out-of-sample (%)	**0.0223**	1.1341	**0.0322**	3.5732	**0.0204**	1.1341	**0.0092**	1.1341
MCSE	0.0000		0.0001		0.0000		0.0001	
LB	0.0184		0.0279		0.0169		0.0090	
UB	0.0263		0.0374		0.0248		0.0105	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

Table 4.4 (cont.)

	SAV		AS		Indirect GARCH		Adaptive	
Approach	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
5% VaR								
$\boldsymbol{\beta}_1$	**0.0209**	**0.0438**	**0.0364**	**0.0328**	**0.0273**	0.0207	**0.3179**	**0.3307**
LB	0.0107	0.0230	0.0268	0.0201	0.0108	-0.0016	0.2785	0.2537
UB	0.0350	0.0646	0.0498	0.0455	0.0488	0.0430	0.3642	0.4077
$\boldsymbol{\beta}_2$	**0.9521**	**0.9554**	**0.9491**	**0.9551**	**0.9532**	**0.9642**		
LB	0.9344	0.9305	0.9363	0.9359	0.9335	0.9556		
UB	0.9649	0.9803	0.9593	0.9743	0.9681	0.9728		
$\boldsymbol{\beta}_3$	**0.0726**	**0.0810**	**-0.0151**	-0.0223	**0.0925**	0.0756		
LB	0.0566	0.0465	-0.0302	-0.0539	0.0672	-0.0479		
UB	0.0942	0.1155	-0.0005	0.0093	0.1262	0.1991		
$\boldsymbol{\beta}_4$			**0.1394**	**0.1354**				
LB			0.1152	0.0850				
UB			0.1661	0.1858				
$\boldsymbol{\sigma}$	**0.1151**		**0.1120**		**0.1153**		**0.1142**	
LB	0.1109		0.1081		0.1111		0.1101	
UB	0.1194		0.1162		0.1196		0.1183	
RQ	**332.69**	379.62	**323.91**	370.02	**333.20**	380.25	**330.12**	374.50
MCSE	0.0032		0.0043		0.0029		0.0046	
LB	332.50		323.69		333.01		330.05	
UB	333.09		324.33		333.61		330.34	
Hits in-sample (%)	**0.0003**	0.0360	**0.0003**	0.0360	**0.0006**	0.0673	**-0.0028**	-0.1204
MCSE	0		0		0		0	
LB	-0.0040		-0.0044		-0.0030		-0.0033	
UB	0.0039		0.0046		0.0043		-0.0023	
Hits out-of-sample (%)	**0.0416**	2.3171	**0.0465**	4.1463	**0.0309**	2.3171	**0.0043**	-0.7317
MCSE	0.0001		0.0001		0.0001		0.0001	
LB	0.0321		0.0385		0.0242		0.0006	
UB	0.0543		0.0558		0.0369		0.0069	

Note: Significant coefficients at 5% and smaller RQ/Hits in-sample/Hits out-of-sample by the Bayesian approach formatted in bold. The performance of CAViaR models is better using the Bayesian approach than the Engle and Manganelli's approach based on Hits in-sample and Hits out-of-sample criteria.

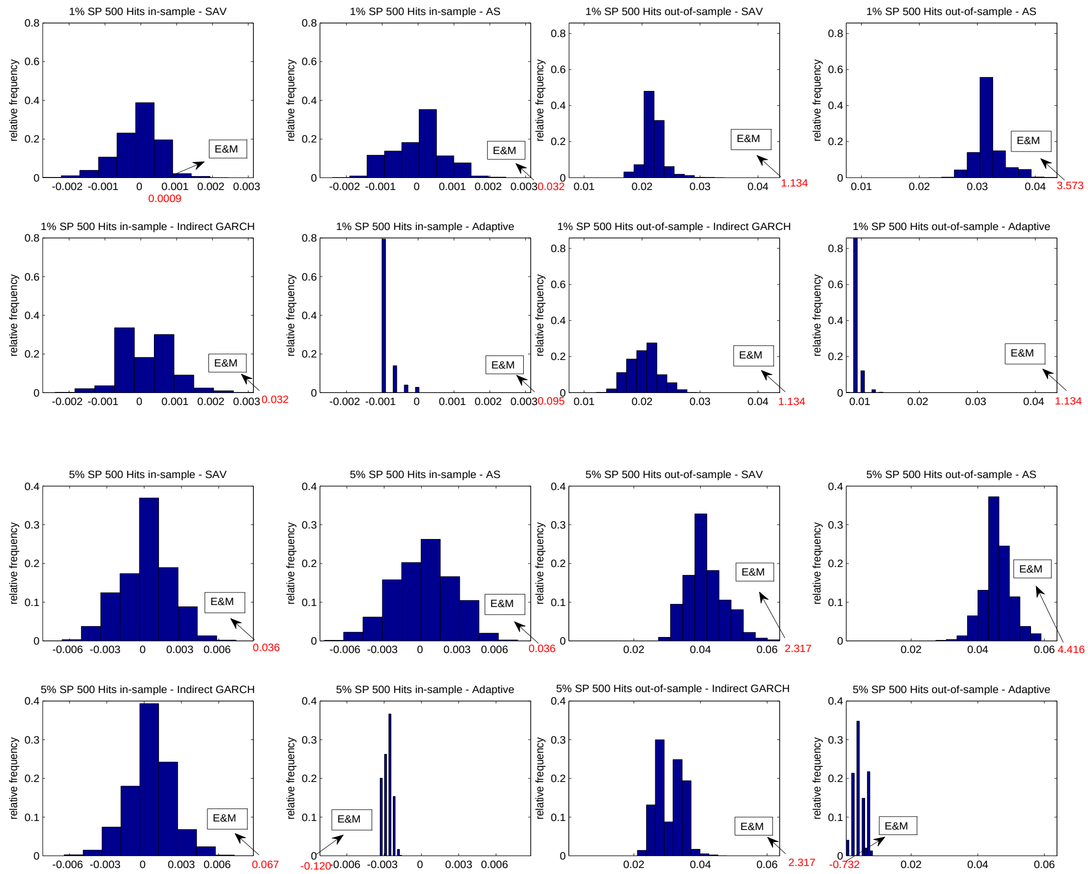
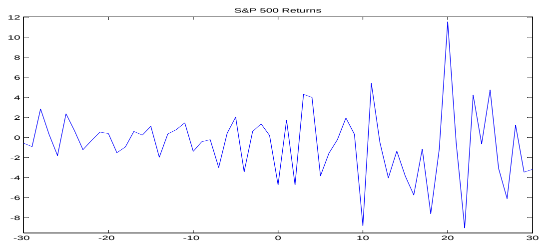
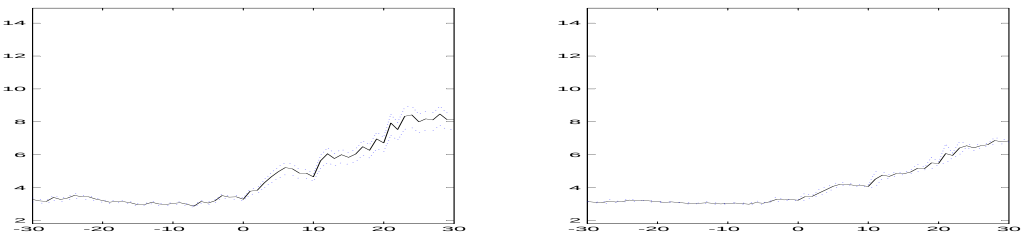
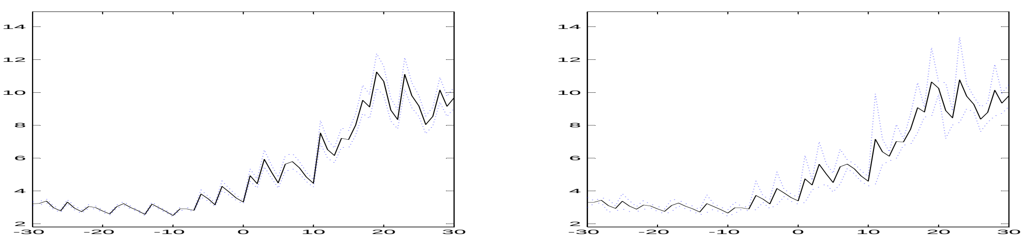


Figure 4.10: Relative Frequency of Hits in-sample and Hits out-of-sample for S&P 500. Hits in-sample are centered around the theoretical value, 0, compared to the Engle and Manganelli's point estimates that are not in the scale of the box except for 1% Hits in-sample using SAV model. Although almost all Hits out-of-sample are not centered around 0, the means are closer to zero compared to Engle and Manganelli's point estimates that are not in the scale of the box.

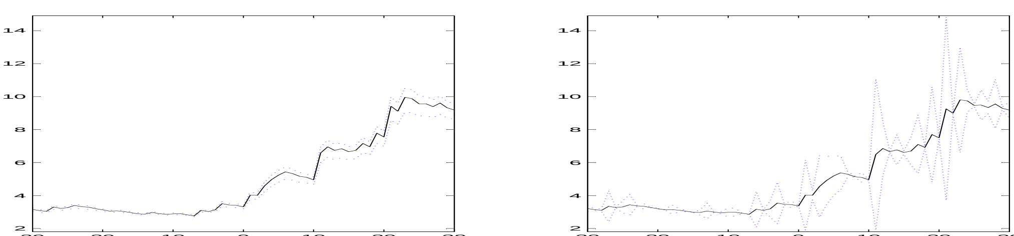




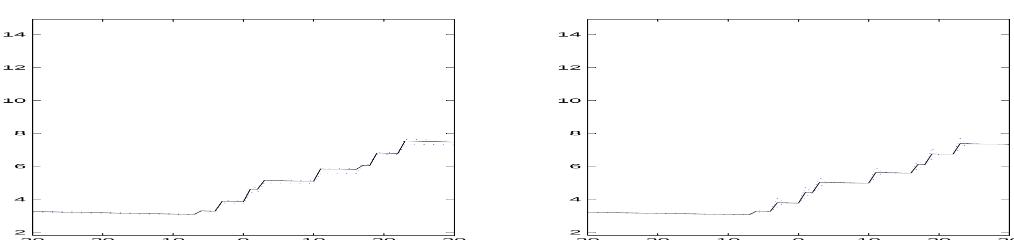
(a) 1% S&P 500 VaR-SAV



(b) 1% S&P 500 VaR-AS



(c)1% S&P 500 VaR-indirect GARCH



(d)1% S&P 500 VaR-Adaptive

Figure 4.11: Returns and 1% estimated VaR plots for S&P 500 index for the 30 days before and after the cutoff. The left plot with 95 % pointwise credible bands is based the Bayesian approach, and the right plot with 95% pointwise confidence bands is based on Engle and Manganelli's approach.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*, 203–228.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Engle, R. F. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*, 367–381.

Flegal, J. M., M. Haran, and G. L. Jones (2008). Markov chain monte carlo: Can we. trust the third significant figure? *Statistical Science 23*, 250–260.

Gelman, A., G. O. Roberts, and W. R. Gilks (1996). Efficient metropolis jumping rules. *Bayesian Statistics 5*, 599607.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–472.

Gutenbrunner, C. and J. Jureckova (1992). Regression quantile and regression rank score process in the linear model and derived statistics. *Annals of Statistics 20*, 305330.

He, X., W. K. Fung, and Z. Y. Zhu (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika 89*, 579–590.

He, X. and Q. M. Shao (1996). A general bahadur representation of m-estimators and its application to linear regression with nonstochatic designs. *Annals of Statistics 24*, 2608–2630.

Hendricks, W. and R. Koenker (1992). Hierarchical spline models for conditional quantiles and demand for electricity. *Journal of American Statistical Association 87*, 58–68.

Hull, J. and A. White (1998). Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives 5*, 9–19.

Keystone, E. C., A. F. Kavanaugh, J. T. Sharp, H. Tannenbaum, Y. Hua, L. S. Teoh, S. A. Fischkoff, and E. K. Chartash (2004). Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis & Rheumatism 50*, 1400–1411.

Koenker, R. (2005). *Quantile Regression* (1st ed.). New York: Cambridge University Press.

Koenker, R. (2010). Rank tests for heterogeneous treatment effects with covariates. See http://www.econ.uiuc.edu/ roger/research/ranks/qte.pdf.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

Kottas, A. and M. Krnjajić (2009). Bayesian nonparametric modeling in quantile regression. *Scandinavian Journal of Statistics 36*, 297–319.

Kremer, J. M., H. K. Genant, L. W. Moreland, A. S. Russell, P. Emery, C. Abud-Mendoza, J. Szechinski, T. Li, Z. Ge, J. Becker, and R. Westhovens (2006). Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis. a randomized trial. *Annals of Internal Medicine 144*, 865–876.

Lehmann, E. L. (1983). *Theory of Point Estimation* (1st ed.). New York: Wiley.

Lipsky, P. E., D. V. D. Heijde, E. W. S. Clair, D. E. Furst, F. C. Breedveld, J. R. Kalden, J. S. Smolen, M. Weisman, P. Emery, M. Feldmann, G. R. Harriman, and R. N. Maini (2000). Infliximab and methotrexate in the treatment of rheumatoid arthritis. *New England Journal of Medicine 343*, 1594–1602.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics 21*, 10871091.

O'Brien, P. C. (1998). Comparing two samples: Extensions of the t, rank-sum, and log-rank tests. *Journal of the American Statistical Association 83*, 52–61.

Owen, A. B. (2001). *Empirical Likelihood*. London: Chapman & Hall.

Serfling, R. J. (2002). *Approximation Theorems of Mathematical Statistics*. Wiley-Interscience.

Sharp, J. T., M. D. Lidsky, L. C. Collins, and J. Moreland (1971). Methods of scoring the progression of radiologic changes in rheumatoid arthritis. correlation of radiologic, clinical and laboratory abnormalities. *Arthritis & Rheumatism 14*, 706–720.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Annals of Statistics 2*, 676–693.

Taylor, J. W. (1999). A quantile regression approach to estimating the distribution of multiperiod returns. *Journal of Derivatives 7*, 64–78.

van der Heijde, D. (2000). How to read radiographs according to the sharp/van der heijde method. *Journal of Rheumatol 27*, 261–263.

van der Heijde, D., L. Klareskog, V. Rodriguez-Valverde, C. Codreanu, H. Bolosiu, J. Melo-Gomes, J. Tornero-Molina, J. Wajdula, R. Pedersen, S. Fatenejad, and T. S. Investigators (2006). Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis. *Arthritis Rheum 54*, 1063–1074.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*, 437–447.

Yu, K. and J. Stander (2007). Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics 137*, 260–276.

Vita

Ya-Hui Hsu was born in Tainan, Taiwan on October 4, 1980, to Mei-Chen Tung (Mother) and Fu-Yung Hsu (Father). She graduated from National Taiwan University, Taipei in 2002 with a Bachelor of Science in Mathematics. She further completed a Master of Business Administration degree in the same university in 2004. At the University of Illinois at Urbana-Champaign, she completed Master of Science degrees in Statistics and Finance in 2008. On completion of her Doctor of Philosophy in Statistics from University of Illinois at Urbana-Champaign, she will start to work as a Statistician at Abbott Laboratories, a healthcare company based in Abbott Park, Illinois.